



US006340583B1

(12) **United States Patent**
Yan et al.

(10) Patent No.: **US 6,340,583 B1**
(45) Date of Patent: **Jan. 22, 2002**

(54) **ISOLATED HUMAN KINASE PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES THEREOF**

(75) Inventors: Chunhua Yan, Boyds; Karen A. Ketchum, Germantown; Valentina Di Francesco, Rockville; Ellen M. Beasley, Darnestown, all of MD (US)

(73) Assignee: PE Corporation (NY), Norwalk, CT (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: 09/813,817

(22) Filed: Mar. 22, 2001

(51) Int. Cl.⁷ C12N 9/12; C12N 1/20; C12N 15/00; C12N 5/00; C07H 21/04

(52) U.S. Cl. 435/194; 435/320.1; 435/252.3; 435/325; 536/23.2

(58) Field of Search 435/194, 252.3, 435/325, 320.1; 536/23.2

(56)

References Cited

PUBLICATIONS

GenEmbl Database, Accession No. D45906, Feb. 1999.*

Sambrook et al., Molecular Cloning Manual, 2nd edition, Cold Spring Harbor Laboratory Press, 1989.*

* cited by examiner

Primary Examiner—Rebecca E. Prouty

Assistant Examiner—M. Monshipouri

(74) Attorney, Agent, or Firm—Celera Genomics; Robert A. Millman; Justin D. Karjala

(57) **ABSTRACT**

The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the kinase peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the kinase peptides, and methods of identifying modulators of the kinase peptides.

9 Claims, 41 Drawing Sheets

1 CCCAGGGCGC CGTAGGCGGT GCATCCCGTT CGCGCCTGGG GCTGTGGTCT
51 TCCCGCGCCT GAGGCGGCGG CGGCAGGAGC TGAGGGAGT TGTAGGGAAC
101 TGAGGGGAGC TGCTGTGTCC CCCGCCTCCT CCTCCCCATT TCCCGCGCTCC
151 CGGGACCATG TCCCGCGCTGG CGGGTGAAGA TGTCTGGAGG TGTCCAGGCT
201 GTGGGGACCA CATTGCTCCA AGCCAGATAT GGTACAGGAC TGTCAACGAA
251 ACCTGGCACG GCTCTTGCTT CCGGTGAAAG TGATGCGCAG CCTGGACCAC
301 CCCAATGTGC TCAAGTTCAT TGGTGTGCTG TACAAGGATA AGAAGCTGAA
351 CCTGCTGACA GAGTACATTG AGGGGGGCAC ACTGAAGGAC TTTCTGCGCA
401 GTATGGATCC GTTCCCTGG CAGCAGAAGG TCAGGTTTGC CAAAGGAATC
451 GCCTCCGGAA TGGACAAGAC TGTGGTGGTG GCAGACTTTG GGCTGTCAAG
501 GCTCATAGTG GAAGAGAGGA AAAGGGCCCC CATGGAGAAG GCCACCCACCA
551 AGAAACGCAC CTTGCGCAAG AACGACCGCA AGAAGCGCTA CACGGTGGTG
601 GGAAACCCCT ACTGGATGGC CCCTGAGATG CTGAACGGAA AGAGCTATGA
651 TGAGACGGTG GATATCTTCT CTTTGGGAT CGTTCTCTGT GAGATCATTG
701 GGCAGGTGTA TGCAGATCCT GACTGCCTTC CCCGAACACT GGACTTTGGC
751 CTCAACGTGA AGCTTTCTG GGAGAAGTTT GTTCCCACAG ATTGTCCCCC
801 GGCCCTCTTC CCGCTGGCCG CCATCTGCTG CAGACTGGAG CCTGAGAGCA
851 GACCAGCATT CTCGAAATTG GAGGACTCCT TTGAGGGCCT CTCCTGTAC
901 CTGGGGGAGC TGGGCATCCC GCTGCCTGCA GAGCTGGAGG AGTTGGACCA
951 CACTGTGAGC ATGCACTACG GCCTGACCCCG GGACTCACCT CCCTAGCCCT
1001 GGCCCAGCCC CCTGCAGGGG GGTGTTCTAC AGCCAGCATT GCCCCCTCTGT
1051 GCCCCATTCC TGCTGTGAGC AGGGCCGTCC GGGCTTCTG TGGATTGGCG
1101 GAATGTTTAG AAGCAGAACAA AACCATTCCCT ATTACCTCCC CAGGAGGCAA
1151 GTGGGCAGCAG CACCAGGGAA ATGTATCTCC ACAGGTTCTG GGGCTAGTT
1201 ACTGTCTGTA AATCCAATAC TTGCTGAAA GCTGTGAAGA AGAAAAAAAC
1251 CCCTGGCCTT TGGGCCAGGA GGAATCTGTT ACTCGAATCC ACCCAGGAAC
1301 TCCCTGGCAG TGGATTGTGG GAGGCTCTTG CTTACACTAA TCAGCGTGAC
1351 CTGGACCTGC TGGGCAGGAT CCCAGGGTGA ACCTGCCTGT GAACCTGTGAA
1401 GTCACTAGTC CAGCTGGGTG CAGGAGGACT TCAAGTGTGT GGACGAAAGA
1451 AAGACTGATG GCTCAAAGGG TGTAAAAAG TCAGTGATGC TCCCCCTTTG
1501 TACTCCAGAT CCTGCTTTC CTGGAGCAAG GTTGAGGGAG TAGGTTTGA
1551 AGAGTCCCTT AATATGTGGT GGAACAGGCC AGGAGTTAGA GAAAGGGCTG
1601 GCTTCTGTGTT ACCTGCTCAC TGGCTCTAGC CAGCCCAGGG ACCACATCAA
1651 TGTGAGAGGA AGCCTCCACC TCATGTTTTC AAACCTTAAATA CTGGAGACTG
1701 GCTGAGAACT TACGGACAAC ATCCCTTCTG TCTGAAACAA ACAGTCACAA
1751 GCACAGGAAG AGGCTGGGGG ACTAGAAAAGA GGCCCTGCC TCTAGAAAGC
1801 TCAGATCTTG GCTTCTGTAA CTCAACTCG GGTGGGCTCC TTAGTCAGAT
1851 GCCTAAAACA TTTTGCCTAA AGCTCGATGG GTTCTGGAGG ACAGTGTGGC
1901 TTGTCACAGG CCTAGACTCT GAGGGAGGGG AGTGGGAGTC TCAGCAATCT
1951 CTTGGCTTGC GCTTCATGGC AACCACTGCT CACCCCTCAA CATGCCTGGT
2001 TTAGGCAGCA GCTTGGCTG GGAAGAGGTG GTGGCAGAGT CTCAAAGCTG
2051 AGATGCTGAG AGAGATAGCT CCCTGAGCTG GGCCATCTGA CTTCTACCTC
2101 CCATGTTTGC TCTCCCAACT CATTAGCTCC TGGGCAGCAT CCTCCTGAGC
2151 CACATGTGCA GGTACTGGAA AACCTCCATC TTGGCTCCCA GAGCTCTAGG
2201 AACTCTTCAT CACAACCTAGA TTTGCCTCTT CTAAGTGTCT ATGAGCTTGC
2251 ACCATATTAA ATAATTGGG AATGGGTTTGG GGGTATTAAGA AAAAAAAAAA
2301 AAAAAAAAAA AAAAAAAAAA (SEQ ID NO:1)

FIG. 1A

FEATURES:

5'UTR: 1-228
 Start Codon: 229
 Stop Codon: 994
 3'UTR: 997

Homologous proteins:

Top 10 BLAST Hits

			Score	E
CRA	1000682328847	/altid=gi 8051618 /def=ref NP_057952.1 LIM d...	485	e-136
CRA	18000005015874	/altid=gi 5031869 /def=ref NP_005560.1 LIM ...	485	e-136
CRA	88000001156379	/altid=gi 7434382 /def=pir JC5814 LIM motif...	469	e-131
CRA	88000001156378	/altid=gi 7434381 /def=pir JC5813 LIM motif...	469	e-131
CRA	18000005154371	/altid=gi 7428032 /def=pir JE0240 LIM kinas...	469	e-131
CRA	18000005126937	/altid=gi 6754550 /def=ref NP_034848.1 LIM ...	469	e-131
CRA	18000005127186	/altid=gi 2804562 /def=dbj BAÄ24491.1 (AB00...	469	e-131
CRA	18000005127185	/altid=gi 2804553 /def=dbj BAA24489.1 (AB00...	469	e-131
CRA	18000005004416	/altid=gi 2143830 /def=pir I78847 LIM motif...	468	e-131
CRA	18000005004415	/altid=gi 1708825 /def=sp P53670 LIK2_RAT LI...	468	e-131

BLAST dbEST hits:

		Score	E
gi	10950740 /dataset=dbest /taxon=96...	1049	0.0
gi	10156485 /dataset=dbest /taxon=96...	975	0.0
gi	5421647 /dataset=dbest /taxon=9606 ...	952	0.0
gi	10895718 /dataset=dbest /taxon=96...	757	0.0
gi	13043102 /dataset=dbest /taxon=960...	714	0.0
gi	519615 /dataset=dbest /taxon=9606 /...	531	e-149
gi	11002869 /dataset=dbest /taxon=96...	511	e-143

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:From BLAST dbEST hits:

gi|10950740 teratocarcinoma
 gi|10156485 ovary
 gi|5421647 testis
 gi|10895718 nervous_normal
 gi|13043102 bladder
 gi|519615 infant brain
 gi|11002869 thyroid gland

From tissue screening panels:

Fetal whole brain

FIG.1B

1 MVQDCQRNLA RLLLKVVMR SLDHPNVLKF IGVLYKDKKL NLLTEYIEGG
51 TLKDFLRSMD PFPWQQKVRF AKGIASGMDK TVVADFGLS RLIVEERKRA
101 PMEKATTKKR TLRKNDRKRR YTVVGNPYWM APEMPLNGKSY DETVDIFSFG
151 IVLCEIIGQV YADPDCLPRT LDFGLNVKLF WEKFVPTDCP PAFFPLAAC
201 CRLEPESRPA FSKLEDSFEA LSLYLGEGLGI PLPAELEELD HTVSMQYGLT
251 RDSPP (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
cAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1 108-111 KKRT

2 119-122 KRYT

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 4

1 51-53 TLK

2 106-108 TTK

3 107-109 TKK

4 111-113 TLR

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 4

1 51-54 TLKD

2 76-79 SGMD

3 139-142 SYDE

4 212-215 SKLE

[4] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 4

1 73-78 GIASGM

FIG.2A

2 77-82 GMDKTV

3 150-155 GIVLCE

4 158-163 GQVYAD

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	142	162	0.872	Putative
2	184	204	0.652	Putative

BLAST Alignment to Top Hit:

```
>CRA|1000682328847 /altid=gi|8051618 /def=ref|NP_057952.1| LIM
domain kinase 2 isoform 2b [Homo sapiens] /org=Homo
sapiens /taxon=9606 /dataset=nraa /length=617
Length = 617
```

Score = 485 bits (1235). Expect = e-136

Identities = 241/265 (90%). Positives = 241/265 (90%). Gaps = 22/265 (8%)

Query: 13 LLPVKVMRSLDHPNVLKFIGVLYKDKKLNLTEYIEGGTLKDFLRSMDPFPWQQKVRF 72
 Sbjct: 353 LTEVKVMRSLDHPNVLKFIGVLYKDKKLNLTEYIEGGTLKDFLRSMDPFPWQQKVRF 412

Query: 73 GIASGM-----DKTVVADFGLSRLIVEERKRAPMEKATTKKR 110
 Sbjct: 413 GIASGMAYLHSMCIIHRDLNSHNCLIKLDKTVVADFGLSRLIVEERKRAPMEKATTKKR 472

Query: 111 TLRKNDRKRYTVVGNPYWMAPEMLNGKSYDETVDIFSFGIVLCEIIIGQVYADPDCLPRT 170
 Sbjct: 473 TLRKNDRKRYTVVGNPYWMAPEMLNGKSYDETVDIFSFGIVLCEIIIGQVYADPDCLPRT 532

Query: 171 LDFGLNVKLFWEKFVPTDCPPAFFPLAACCRLEPESRPAFSKLEDSFEALSLYLGELGI 230
 Sbjct: 533 LDFGLNVKLFWEKFVPTDCPPAFFPLAACCRLEPESRPAFSKLEDSFEALSLYLGELGI 592

Query: 231 PLPAELEELDHTVSMQYGLTRDSPP 255
 Sbjct: 593 PLPAELEELDHTVSMQYGLTRDSPP 617 (SEQ ID NO:4)

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
PF00069	Eukaryotic protein kinase domain	100.1	1.1e-26	2
CE00031	CE00031 VEGFR	4.9	0.14	1
CE00204	CE00204 FIBROBLAST_GROWTH_RECECTOR	4.7	1	1
CE00359	E00359 bone_morphogenetic_protein_receptor	1.8	7.9	1
CE00022	CE00022 MAGUK_subfamily_d	1.5	2.5	1
CE00287	CE00287 Eph_orphan_receptor	-48.4	3.8e-05	1
CE00292	CE00292 PTK_membrane_span	-61.8	2.1e-05	1

FIG.2B

CE00291	CE00291	PTK_fgf_receptor	-113.0	0.027	1
CE00286	E00286	PTK_EGF_receptor	-125.1	0.0021	1
CE00290	CE00290	PTK_Trk_family	-151.3	6.5e-05	1
CE00288	CE00288	PTK_Insulin_receptor	-210.4	0.014	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00069	1/2	16	79 ..	41	105 ..	52.1	2.3e-13
CE00022	1/1	124	153 ..	187	216 ..	1.5	2.5
PF00069	2/2	81	156 ..	129	182 ..	48.0	3.1e-12
CE00031	1/1	129	156 ..	1114	1141 ..	4.9	0.14
CE00204	1/1	129	156 ..	705	732 ..	4.7	1
CE00359	1/1	79	157 ..	287	356 ..	1.8	7.9
CE00290	1/1	9	218 ..	1	282 []	-151.3	6.5e-05
CE00287	1/1	1	218 [.	1	260 []	-48.4	3.8e-05
CE00291	1/1	1	218 [.	1	285 []	-113.0	0.027
CE00292	1/1	1	218 [.	1	288 []	-61.8	2.1e-05
CE00288	1/1	1	218 [.	1	269 []	-210.4	0.014
CE00286	1/1	6	218 ..	1	263 []	-125.1	0.0021

FIG.2C

1 TCATCCTTGC GCAGGGGCCA TGCTAACCTT CTGTGTCTCA GTCCAATTT
51 AATGTATGTG CTGCTGAAGC GAGAGTACCA GAGGTTTTTG TGATGGCAGT
101 GACTTGAAC TATTAAAAG ATAAGGAGGA GCCAGTGAGG GAGAGGGGTG
151 CTGTAAGAT AACTAAAAGT GCACCTCTTC TAAGAAGTAA GATGGAATGG
201 GATCCAGAAC AGGGGTGTCA TACCGAGTAG CCCAGCCTT GTTCCGTGGA
251 CACTGGGGAG TCTAACCCAG AGCTGAGATA GCTTGAGTG TGATGAGCC
301 AGCTGAGTAC AGCAGATAGG GAAAAGAAC CAAAAATCTG AAGTAGGGCT
351 GGGGTGAAGG ACAGGGAAGG GCTAGAGAGA CATTGGAAA GTGAAACCAG
401 GTGGATATGA GAGGAGAGAG TAGAGGGTCT TGATTTGGG TCTTCATGC
451 TTAACCCAAA GCAGGTACTA AAGTATGTGT TGATTGAATG TCTTGGGTT
501 TCTCAAGACT GGAGAAAGCA GGGCAAGCTC TGGAGGGTAT GGCAATAACA
551 AGTTATCTTG AATATCCTCA TGGTGGAAAG TCCTGATCCT GTTGAATT
601 TGGAAATAGA AATCATTCAAG AGCCAAGAGA TTGAATTGTT GAGTAAGTGG
651 GTGGTCAGGT TACAGACTTA ATTTGGGTT AAAAAGTAAA ACAAGAAC
701 AAGGTGTGGC TCTAAAATAA TGAGATGTGC TGGGGGTGG GCATGGCAGC
751 TCATAAAACTG ACCCTGAAAG CTCTTACATG TAAGAGTTCC AAAAATATT
801 CCAAAACTTG GAAGATTCA TTGGATGTTT GTGTTCTTA AAATCTCTCA
851 CTAATTCTT GTCTGTCCA CTGTCGTAA CCCAACCTGG GATTGGTTG
901 AGTGAGTCTC TCAGACTTTC TGCTTGGAG TTTGTGAGAG AGATGGCATA
951 CTCTGTGACC ACTGTCACCC TAAAACAAA AAGGCCCTC TTGACAAGGA
1001 GTCTGAGGAT TTTAGACCCA GGAAGAAAGA GTGATGGCA TATATATATC
1051 CTATTACTGA GGCATGAGAA GAGTGGAAATG GGTGGGTTGA GGTGGTGT
1101 TAAGGCCTCT TGCCAGCTT TTTAACTCTT CTCTGGGAA CGAGGGGGAC
1151 AACTGTGTAC ATTGGCTGCT CCAGAATGAT GTTGAGCAAT CTTGAAGTGC
1201 CAGGAGCTGT GCTTTGTCTA TTCACTGGCCC CTGTGCCTGT GAAACAGGGT
1251 TCGGTGACTG TCACTGTGCC TGTGGCAGTC TGTAGTTACC CAGAGAGAAC
1301 AAAGCTGCAT ACACAGAGCG CACAAGGGAG TCTTGTAAACA ACCTGTCTC
1351 GCTTCTAGG GCTGAGTCAG GTACCAACAGC TTGATCTCAG CTGTCCTCTT
1401 TATTTCAAGA AGTTGACATC TGAGCCATAC CAGGAGTATT GTATTTGTT
1451 TGAGGCCTCT CTTTTGGAG GAACATGGAC CGACTCTGTG CTTTTGTCTA
1501 TGCTGGTCTC TGAGCTCACA CAACCCCTCA CCCTCCTTC TCAGCCAGTG
1551 ATAGGTAAGT CTTCCCTATC TTGCAAGGCT CAGCTCAAGT GTCAGCTTCC
1601 TCTACAAAGA CTTTCTGGT TCCCCTCATT GGAGTGAACA AGAGTTGACA
1651 TGGTAGAATG GAAAGAGCAG AAGCTTACA ATGAGCCAGA CCTGAGTATG
1701 AATGCTAGAT CCACCACTTA GCTAGTCAC CCTGCCCCCT GCCTCAAGTT
1751 TTAATTTCC TATCCATTAA GTGAATATAA TAATACCTGT GTCACAGGGAT
1801 TATTTTGAGA ATTAATGAG ATTAGGTCTA TGAAAGCACC TAGCAGAGTT
1851 CTTGGCATAT AGGAGGCATT CATTAAATAT TTGTTCTTCC CCTTTTATAAC
1901 CCATTACTTT TCTTTTCTG AACTAAAATA ATACTGGTT CTATCTCTGA
1951 AATAACATCC AAGTAAAAAA TCAACAAACAT GAAAGAGCAG TTCTTTTCCA
2001 GTGGATTTGC TTCTTAAGGA GCAGAGATTA TGTAATCTAA CAGCCTCCAA
2051 CATAACAAAGA GCTTTGTATC TAGAACAGGG GTCCCCAGCC CCTGGACCGC
2101 CAACTGGTAC GGGTCTGTAG CCTGTTAGGA ACCAGGCTGC ACAGCAGGAG
2151 GTGAGCGGGCG GGCCAGTGGAG CATTGCTGCC TGAGCTCTGC CTCCGTCTCAG
2201 ATCACTGGTG GCATTAGATT CTCATAGGAG TGTGAACCCCT ATTGTGAAC
2251 GCACATGCA AGGATCTGGG TTGCACTGCTC CTTATGAGAA TCTCACTAAT
2301 GGCTGATGAT CTGAGTTGGA ACAGTTGAT ACCAAAACCA TCCCCCCGCC
2351 CCCCCAACCC CAGCCTAGGG TCCGTGGAAA AATTGGCCCC TGGTGCCAAA
2401 AAGGTTGAGG ACTGCTGATC TAGAGGACCA ATTATTCAA TGTTGGTTGA
2451 GTAAATGAGC TCTGGATTA GGTGATGGAA AAATCTGAAA AACAGGGCT

FIG. 3-1

2501	TTTGAGGAAT	AGGAAAAGGC	AGTAACATGT	TTAACCCAGA	GAGAAGTTTC
2551	TGGCTGTTGG	CTGGGAATAG	TCATAGGAAG	GGCTGACACT	GAAAAGAAGG
2601	AGATTGTGTT	CGTTTCTTCT	TCTCAGAGCT	ATAAGCAAAG	GCTGAAAGTT
2651	CTAGAAAAAG	GCAAGTTTTG	TTTCAGTAGA	AAAAGGATA	ATCAGAACCA
2701	TTTTTAGAAA	ATGGAATGAG	ACTACTTTG	AGGCCATGAG	TTCCCTTGTC
2751	CTGGAGAGAT	GAGCAGAGGT	TGGACAAGTG	CTTACCAAGAG	ATCTTGTGGA
2801	GGCAGAAACT	GTGCATCTAG	CAGAGCATTG	GCCTAACCCCT	TTCAAATGAG
2851	ATGCTGTTAA	CTCAGTCTTA	TTCTACATGG	TAGGAATCCT	GTCCCCTTGC
2901	CTCCTGCTAC	TTGGGGCCTC	TCAACCTTT	GGTTTGTGT	GCAGGTGAAG
2951	ATGTCTGGAG	GTGTCCAGGC	TGTGGGGACC	ACATTGCTCC	AAGCCAGATA
3001	TGGTACAGGA	CTGTCAACGA	AACCTGGCAC	GGCTCTTGCT	TCCGGTAGGT
3051	GGGCCTATCC	TCCCATCTT	ACCACTGTAC	TATGGGCCAA	GCACTATTT
3101	ATGTTCTGAT	GGAAAACACA	GAAACAAGCT	TCTGAGTTGA	GAATTTCAAT
3151	CTTAGGGTGG	GGAAAAGGAAT	GTACCAAGGA	AGAGCTCATG	ACCAAACCTC
3201	AAGTGTGGCC	CCCCTGAACC	CAGGTTAAAT	TGGAAGAGCC	ATAAATGGGC
3251	CAGCTGGAGG	CAGGGTGGGG	GGATGAGAGG	AGCCCTTCC	AGGGTTGTCC
3301	CATATCCCTC	ACTTTATGGG	TGAGGAAACT	GAGGCCAGG	AAGAGTGA
3351	TTCCCTGTGGC	TGCACTACAG	ATTATGCAGG	TACTTCAAGA	GTGTTTGT
3401	TTCTTATTTT	ATTTTATTTT	ATTTTATTTT	ATTTTATTTT	ATTTTATGAG
3451	AGGGATTCTT	GCTGTTGCC	AGGCTGGAGT	GCAGTGGTGC	AATCTCGGCT
3501	CACTGCAATC	TCTGCCTGCT	GGGTTCAAGT	GATTTTCTG	CCTTAGCTTC
3551	CTGAGTAGCT	GAGATGACAG	GCACCTGCCA	CCATGCGCAG	CTAAFTTTG
3601	TATTTTAGTG	GAGACGGGGG	TTTCAACATG	TTGGTCAGGC	TGGTCTTGA
3651	CTCCTGACCT	CAAATGATGC	ACCCACCTCG	ACCTCCAAA	GTGCTGGAA
3701	TACAGGCGTG	AACCACTGTG	CCCAGCCAAG	AGTTGTTTTT	AGTGTGGTTG
3751	GCAGAGCCAG	CTCTCCCTTC	ACCACAGGAT	GCCTCCCTAG	GTTCCTACTT
3801	TTTGTACTA	GCTTTTATT	TAGCTATATT	ATTATTATT	TTATTATTAT
3851	TATTATTATT	ATTATTGAGA	CAGAGTCTCG	CTCTGTCGCC	CAGGCTGGTG
3901	TACAGTGGTG	CGATCCCAGG	CTCACTGCAA	CCTCTGCC	CCGAGTTCAA
3951	GCAGTTCTCC	TGCTTCAGCC	CCCCGAGTAG	GTGGGACTAC	AGGCGCCTGC
4001	CACCACACCC	GGCTAATTTT	TGTATTTTA	GTAGAGACGG	GGTTTCACCT
4051	TGTTGACCAG	GCTGGTCTGG	AGCTCCTGAC	CTCAGGTAAG	TGCTAGAAC
4101	ACAGGCGTGA	ACCACTGCGC	CCAGCCAAGA	GTTGTTTTA	GTGTGGTTGG
4151	CAGAGCCAGC	TCTCCTCAC	CACAGGTTGC	CTCCCTAGGT	TCCTACTTT
4201	TGTTACTAGC	TTTATTATAG	CTACATTATT	ATTATTATTG	TTATTATTAT
4251	TGAGACAGAG	TCTCGCTCTG	TCGCCCAGGC	TGGTGTACAG	TGATGTGATC
4301	TTGGCTCACT	GCAACCTCTG	CCCCCCCAGT	TCAAGCAATT	CTCCTGCTTC
4351	AGCCCCCTA	GTAGGTGGGA	CTCCAGGCAC	CTGCCACCAC	GCCCAGCTAA
4401	TTTTGTATT	TTTAGTAGAG	GCGGGGTTTC	ACCTGTTGG	CCAGGCTGGT
4451	CTCAAACCTCC	TGACCTCAGG	TGATCCGCC	GCCTCGGC	CCAAAATGT
4501	TGGGATTACA	GGCATGAGCC	ACCGCGCC	GCCTATAGCT	ACATTATTT
4551	TGTAGGCAGC	TCAGTTTCTT	AAAAATTATA	CAGACTCAA	ATCAGATTG
4601	TTCCCTGCTGT	CTGAGGCTCA	GTTCCTTCAT	CTGGAAAATG	GATGGAATA
4651	ATCTTGTGAA	GATTGAATGA	AATAATATAT	GCAGTGTATC	CAGTACATGG
4701	TAGACACCCA	GTGAATGGTT	ATTCCTTCCT	CCCATCGGAT	TGGAATTCTC
4751	AAGGGTGGGA	ACTTGTCTT	ATATTCTCA	CAACGTAAAA	TAGTTGAAT
4801	TTGTTGGTGG	AAAGAAGAGC	AGTCCACTCC	AGAGGCTGGA	TGGGCATGCC
4851	TGGCCCCCAA	GGTCTGAAGT	GGTAGGGCTG	TGCCTATATC	CTGAGAATGA
4901	GATAGACTAG	GCAGGCACCT	TGTGCTGTAG	ATTCCAGCTC	CTGCACATAG
4951	CTCTTGTGT	AAAACATCCC	TGTGCTTATA	CCAAGTAATT	GAGTTGACCT

FIG.3-2

5001 TTAAACACTT GCCTCTTCCC TGGGAACCAT ATAGGGGATT GGCCTGGAGA
5051 CGTCTGGCCT CTGGAAGAGT TGGAAAGCAG CCATCATTAT TATCCTTTCC
5101 TTTCAGCTAT AACTCAGAGC TCTCAAGTCT TTTCTGTGGA TCTTATTGCC
5151 TTGGTTCTTGC CCCCTTTAC TCCCAGGGAA GTTGATTCTG TCTTTCTGT
5201 TCCATTTAGT ATGACAGGAG CAGAGAATGT CAGAGCTGTA AGGGACCTTA
5251 TAGTTAAAGC CTTGGCTGG TCCTTTCATT TTATAGCTGG GACTAATAAG
5301 TAACGTAAA ACCCAATGAG TTCACAGATT GGGCTCGCC TTGGCATGTA
5351 ACCCATATGT TCATATTCTT GCTGTTTCC TATGTGTATG AATATTTCT
5401 ATCCAAAATA AGCAGGACAG GGTAGAGCAA GTTAATCTT GGAATTCTG
5451 GATTCTCTTA GAGCTAAAAA ACTTCAGAAC TAGAAGAAC CACCCACTAT
5501 ATGGTATAAC CCATTCAAT CACAGATGAG GCCTGAAACC AAAAAGACTT
5551 GCTCAGGCCA TGGATGACAA GAGCTGGCC TAGCACTGAA CTCTTGGGTC
5601 ATTTGTAGGT CTAGTCAGAT GCTAGCTTGT TAGCTCTGTG CGTGCCTGTG
5651 TGTGTGTGTG TGTGTGTGTG TGTGTGAGAT AGAGACAGAA AGATAACATA
5701 TGTACACAAA TACATAAAGA GGAAGTAGAC ACGTTAGCAT GGTAGATAAG
5751 AGTACAGGCA GGCCAGGCGT GGTGGCTAC GCCTGTAATC CCAGCACTT
5801 GGGAGGCCAA GGCAGGTGGA TCACCTGAGG TCAGGAATTG GAGACCAGCC
5851 TGACCAACAT GGTGAAACCC CATCTCTACT AAATACAGAA AAAAATTAGC
5901 TTGGCATGGT GGCACATGCC TGTAATCCA GCTACTTGGG AAGCTGAAGC
5951 AGGAGAATCG CTTGAATCCG GGAAGCAGAA GTTGCAGTGA GCCGAGATTG
6001 TGCCATTACA GTCTAGCCTG GGCAACAAGA GGGAAACTCC ATCGAAAAAA
6051 AACAAACCA ACCAAGAGTA CAGGCTATGG AATGAGACTA TGGTTTAA
6101 TCCTGGCTTT GCAATTATT AACTAGCCTT AAGTGAATT CTCAGCTTC
6151 AGGCACCAAT CTGTAAAATG AGGATAAGAA TATTACTCAT GCCACATGGT
6201 TGTAGGGAG GATTAATGT GATAACCTAT ATAAAGTGGC TAGCATAGCA
6251 TCTGACATAT AGAAAACCTCT TAATAGGGCC GGACGTGGT GCTTATGCC
6301 GTAATCCTAG CACTCTGGGA GGCGAGGCA GAAGGATCGC TTGAGCCCCAT
6351 GAGCCCAGGA GTTGAGACC AGCCTGGCCA ACATGGCAAA ACTCCACCTC
6401 TACAAAAAAT ACAAAATAT TAGCCAGGCG TGATGGCACA CACCTGTAGT
6451 CCCAGCTACT TGGGAAGCTG AGGAGCGATG ATTACCTGAG CCCAGGGATA
6501 TCAAGGCTGT AGTGAGCTGT GATCATGCCA CTGACTCCA TCCAGCTGG
6551 GGACAGAGTG AAACCCCTGT CTCAAAACAA AACAAATGAA AAAAANAC
6601 CTTAATAATC AGTAACGTG ACTTTATATT ATGTTGTGAG TGTGTGTCTA
6651 TATACACCTA TATGTATACA TTTCTCTTAT TACACATTCA TTGGTGATCT
6701 GATGTGGAGC CCCAGGGATT AAGGGCAACT TTGAACCTACC CTGACACAAAT
6751 CAAGCCAAAT ATCATTCCCG TGGAGGAAGT AGAGTATCTA GGTTCTGTCT
6801 CCTAGTTGCA GCTTACCTT GAGGACAGAG ACTCTAATCC AGCTGTGCTG
6851 AAGGAGCACA TCTCTGACT TCTGAGCTT CCCCTGGTAA ATTCAAACGT
6901 GATGTACGG CGCCCTCAGA TAGAGCCTGG TAATTGCC TGGGGAGAGT
6951 GACTGTCTT TGGATCTAAT TTGACTTTTG CCCAGTTGG AGGAAAATCT
7001 TCAGGGCTAG GAAGGATTGT ATTTGTCTGA CCCCAGAGAT AACCTGGGTT
7051 TTGAGGAACA TGGGGCATCA ACCTGAATGG TCTTGTAAAGA TCTCTCCCAC
7101 GCCAGCTTGC CAGTCTTCT CTGATGAATT TAGAGTACCT GAGTAGTGCA
7151 GGCCTGCTGG GAGGAGGACT CTCCCTCTGT GCTACTCAGA GAAATTCTT
7201 CTTCAAGGCC CCCCTCCAGC CTTGCTCTTA CCCAGCTGGG CTACAGTTAC
7251 AATAAAGGAA ATGACTTTTC TTCTCCCCCTT CCCCCAGTAC CTTTGTCTT
7301 CTAGTCACAG GGTGGGGCTG GATATTGAAT GGAGAAATTG CTGGGGTCCA
7351 TCCTAAACTC CTCCCCCTCAT CTCTCCCTTA CATTACCCCA TTCTCTGTG
7401 TGCAGGCCACA TCCATAATCC TGCCCTGTGTT AGCCTCCGA CAGACCCCTCA
7451 GGTGCCAGG ACAACAGGAA GCTACTAAA GCTGGAACCT CAGACTGTG

FIG.3-3

7501 AATGGAGGCC AGTGACAAAA CTGAAAGTAG CTCTGTCAGT AATTGTGCTG
7551 GTGCGATTAG GCAGCTGCC AGAATCTTT GGATCTCCTG GACATATGGC
7601 TGACTAGTCC TCCCAAGCCT TCCCAACAGG CCTCTTTTTT TTCCCTTTTT
7651 TCTTTCTTT TTTTCTTTC TTTCTTCTT TCTTTTTTT TTTTTTTAG
7701 GCTAGTGAAG TGAATTGTG GGAGTGGAAA AGGAACAAAG AAATCGGTAA
7751 CTGGTAGTGA TCAATTACTT GTAAACACTA TTGTACTTGG ACCAGCCCAG
7801 TAGGCCTTT TAAAAACTCT GAGTTACCTC TCTTCCCTT CCTTGAGCAG
7851 TGCCATTAAT TCTGTATCTG GGGCAATCCT TTCTGATGTT CTCTGGACCT
7901 GGCTCTCTC CCTTAGGAGA GGCCAGGAGA GTAGCCAGAG AGCATGTCT
7951 TTGTAGCTGA GGTAAAGTG TGGAGCTATC AATGGTGACC TGGCCTCTT
8001 GCATGTTAGC AAGCCAGAGG ACCTTGACAA CTTTTTGAT GATTGTCGT
8051 TCACCTCTGAT CAAAGGTGTT TGGCTTAGGA GGAGGGAAAGA AAAGCTACCC
8101 CTATTAGTCT TGATGGCCCC AGCGTGGTCT TCTATTGCTT GACCTGGTTC
8151 CTAGCAGCAT TATCAGAAGG AAAATCCACC GCTCTTAAGG CTCTGGGAA
8201 CTTTCAGGAC TTCTTTCTC AGGATTGCAA ACATAAGACT ATTTGAGCTT
8251 TCACTTTGA AAAGCGGTTA CTAATACCTA TACTCTGGGA AAGGGCTAAT
8301 GCAGATAGAA GACTGTGGTC ACTGCATCAG GCAACAGACC ATTTCCGCTA
8351 AATTTAGTGA CTCCAGGAAG GCCAGTGAAG AAATAACACA CGTAGCAACC
8401 AGAGACTGTG TTGTAATATG TTGGCTGACA GCAGGGTACT TTCTGATG
8451 CTGAAAGCCA CATTCACTT CTCTCCCCTC ATCCCCATCT AAGCAAGCCT
8501 GGTAGAATCA TAATTACAGT AATAGGTACC ACTTATTGAG TACTCTGTG
8551 CAGACACCCCT CCTGAGCATA CGACATGCAT AGCACATTAA ATCCTTACAA
8601 TGACTTAATA AAATGTAGTA CTAGCTTAC CTACTTCGAG AATAGGGAAA
8651 TGGAGGTTAC TTGTTTAAAG TCACAGAGCT AATAGGTAGC ATAGCTGAGA
8701 TTGAACTCA GGCATTCTTA CTCCCTGCCT GCAAGAGTCT CTTGGCATT
8751 TTGAATGCAA GCATATTCT TAACCTCACT GAGGCTCAGT TTCCCTTTAT
8801 ATAATATGGG GTAAGAGGCC CTCACCCCTGC CTGCCACACA CTGGTAGTGT
8851 CAGATAACAT TGAAAGGGTGT TAGTTAAAG GCTTCATGGA CTCTATAATG
8901 TCAACAAAAG TGCTGTTAAC TTCTCTGG GTCTCAGGCT CCTGATGTAG
8951 AGTCAGTGGA GCAACCCCTGC CATCTGCTGT TATGCTGTTG ATGTTGCTG
9001 CACACTTACT AACCTAAACC TTGATTCTG GCTGTGGCCT TCTCCAGAAG
9051 GTGTTTACTC ATTGTCCAG TTATCTTT AGGAAACAGC CAGCCCGTAG
9101 ATCATTAAAGG CTGGCTATTG GACAGGGGGC TGGGGCCTGC CTGACAGAGG
9151 AAGGAAGGGC AGACATCTGG TTCTCTCTC GCCCCTACAA GAGACTCCAG
9201 CCTGACCACA GAGTGGTACT CCTAGGATGT AGCAGCAGCA TATGAGCTTG
9251 AATGTGCCTT AATCCTGCTC TTACTTTGA GAAGAGAGAA CTAAGGACCC
9301 ACAGATGTTT CACAGCTTCT ATAGGAGGCA GAGGTAGAAA AATGGAGAGA
9351 GATGAGGCCA GAGATAGATA ACTGATATTA ATAAACGTT GTATTAAGAA
9401 CCTCACTTAG ATTATCTGAT TCAATCTCA TAATAACCC GCAACCCCCA
9451 CCTTTTTTG AGAACAGGGT TTGCTCTGT TGTCAGGCT ACAGTGCAC
9501 GGTACAATCA TAGTTCACTG CAGTGTCAAC CTCTGAGCT CAAGCAATCC
9551 TCCCACTCTA GCCTTGCAAG CAGTTGGAC TACAGGCCTG CCACCCAC
9601 TTGCCATTCTT TTTTATTTT AAGTAGAAC AAGGCTTAT TAATACTATG
9651 TTGCCAGGC TGGCTTGAA CTCCAGCGAT CCTCTGCC CAGCCTCCCA
9701 AAGTGTGTTGG GATTACGGAA GTAAGCCACT GTGCCTGGCC AGTGCAACCC
9751 CCATTTATA CTAACACAGG AAGGCCAGA AAGGTTGGA GTAACTTGT
9801 CAGGGTCACA CAGATGATAT TTGAACTCAG GTCTCCCTGG CTCCCAAGAG
9851 AGTCTGCTTT CCACTAGGAC TCCCAGGAGA AAAAAAAA AAAAAACAGT
9901 AGACTGGAG ACAGAAAATC TGATTTGAGT CTTAGTTGAG CTAGGCTAAC
9951 TGTGTAACTG TGGGCAAGTT CCTAGCCCC TGTGAGCCTC AGTTTCTTAT

FIG.3-4

10001 CTGTAAAATG TCATAAAAGA AATCCATCTC ATGGAGTAGT TGTGATGATC
10051 AAGGACTCTG AAAACATTAG AATGGTTAA TGTGAAGGAT TAGCAGCAGC
10101 ACATGGCAAC ATTGTGCATC TTATATTAAAC TATCAAATA TATCAAGCGT
10151 CATTGCTAT ATATAAAAGT CATCAAATTA GGCACGTGAG GGGATACGGA
10201 GTTGGCATAAC TAGCCTGGCC TCTTAATTAA TTCATTAATT AGCTTATTAA
10251 TTTTGAGAT AGGTCTTGCT CTATTGCCA GGCTGGAGTG CAGTGGCATG
10301 ATGATAGCTT ACTATAGCCT CAATCTCCA GGCTTAAACA ATCCTCCTGA
10351 GTAGCTGGGA CTACAGGCAC ACACATACCAT GCCCAGCTAA TTTTTTTTAA
10401 ATTTTTGTA GAGACAGGGT CTTGCTCTGT TGCCAGGCT GGTCTCAAAC
10451 TCCTGGGCTC GAGATCCTCC CACCTGGGCC TCACAAAGTG TTGGGATTAC
10501 AGGTATGAGC CACGGCACCT GGCCCTGGTCT CTTAACTGGT TCCCTAAGAC
10551 AGCTGAAAT AGAGAATGTC ATGGAGCATT CCTAACCATG GGCTCCAGCC
10601 TGGCTTTCAT TCTGTTTCTC CCCTGAAACA ACATTCCTT AGTAATATTC
10651 CGAATAACAG CTTCATCAGT CTGTCCTACCG ACCACTCTTC AGGCTTCATC
10701 TTATATGACC TCCCAAACCTG CACTAAGGGT TGTATTAGAG AAAAGTGGAT
10751 AAAGTTCGGA GTCAAGGCTGC TTGAGCTTAA ATGCCAGCTT CACTTACCAAG
10801 CCACCTGACC ATGAGTCAGC TGCTTAACCA TTCTTGCCA CAGTTTCTT
10851 GTCTATGAAA AGGGAAATGG CTCCACCTC AAAAGTTGT TAACATTAAA
10901 TTCAATCATG TATTCAAAGT CCTGAGCAGA ATGTCAGGCC ATGACTGGGA
10951 CTTAACAGAT GTTAGCATTT ATTATTAGTA TCTGTCAGTC TTGAAATGTT
11001 CTCTCCCTT GGCTTCATG ACATTCACACA CTCTCCTGGT TTTCTCTTAC
11051 CTCTCTGGTA ATACCTGTTT GCTTATCCTT CTTTGTCAG CTCTGGGATG
11101 TTACCATTCCTC TTCAAGGCGTG CTGTTTCTC CTTAGGCAGT CTTACACACA
11151 CTCATGACTT CCTCCATTG TCCTCCACAC ACTGATGACC CTAAAATCAG
11201 TATCTCCAGC CAAACCTTT CCACTGAGTT CTAGACCCAT ATGTTGTACT
11251 ATCAACCTGG CTTGTCATT TGAATGTCTT CCAGGCACCT CAGACTCTCT
11301 TCTCTAGACT TTGCTGGACT TTCACTCTTC CCCCTAAAAC TGGCTCTCT
11351 TCCACTGAAA CATGTATGTC ATTGAGAGGC ACCACCATCC ACCCAGTGCC
11401 TAAGCCAGAA ACCTAGGAAT CCTTGATACC TGTTCTCT CATCCTGCAT
11451 ATCCAAGCCT ATCAGTTTA TCTCTAAATT ATATTTGGT AGGTTTACTT
11501 CTTCCTTTT CTCCCACAC CACCCCTGTC CAAGCTACCA TCATCTCACC
11551 TGGATGTCTG CAATAGCCTC ATCTCCACACA GCCACTCTGC ACCCCCCTAAT
11601 CTGTTCTCTA TAGAGCAGTT GGAAGGAGTG ATTTTTGTTG TTTGTTTGT
11651 TTTGTTTGTAG ACAGAGCTC ACTCTGTTCC CCAAGGCTGG AGTGCAGTGG
11701 CACAATTTCG GCTCACTGCA ACTTCTGCCT CCCGGTTTA AGCAATTCTC
11751 CTGCCTCAGC CTCCCAAGTA GCTGGGATTA AGGCACCGGC CCCCATACCC
11801 AGCTAATTTT TATATTTTA GTAGAGATGG GGTTTTGCCA TGTTGGCCAA
11851 GCTAGCTCG AACTCCTGAC CTCAAGTGAT CCACCTGCCT CGGCCTCCCA
11901 AAGTGCTGGG ATTACAGGTG TGAGCCACTG CACCTGGCTG GAAGGAGTGA
11951 TCTTAAAAAA AAAAAAAACA AAAAAAAACT TGACTGTGTC ACTCTGTGTT
12001 GTCTCTCTA CCTTGATAC TTCCACAAC TCCCAGTGTT CTTGGATAAAA
12051 GACCAAAATC CTTAACTTGG CCAGGCGCGG TGGCTCACAC CTATCATCTC
12101 AGCACTTGG GAGGCCGAGG CAGGCAGATC ATGAAGTCAA GAGATTGAGA
12151 CCATCCTGGC CAACATGGTG AAACCCCATC TCTACTAAA ATACAAAAAT
12201 TAGCTGGCTG TGGTGGCGTG TGCGCTGTAGT CCCAGCTACT TGGGAGGCTG
12251 AGGCAGGAGA ATCACTTGAA CCTGGGAGGC AGAGGTTGCA GTGAGCCAG
12301 ATCACGCCAC TGCACCTCAG CCTGGTGACA GAGTAAGACT CCATCTCAA
12351 AAAAAAAACA AAAAAAAATAA TTCCCTTAATT TGGCCTACAG TAGAGCCCTC
12401 CGTAATGTGG CCTCTCTCCA CATCTCCACA ACCTCCTGCT CCCTGACTT
12451 CAGCCTCACC TCTCTCTGG ACAGGCCCTC CTTCTGACAA GGGCTTTGTT

FIG.3-5

12501 CATTCTGCTC CCTCTGCCTA GAATGCCCCC TTACTCTGTT CACTTAACTC
12551 CTGCTTATCG TTTAGATCTT TACCTGGATG GCTCAGAGAA ATATAGAAGT
12601 AATTCCCTCAC CCTGAAAAAT AGGTTAGGTC CCTGTTTAT GTTTTCATAG
12651 ACCCTTCCTT TGAGGCTTTT TTTAAAAAAG TAGTTTAAT CTCACATTTA
12701 TTCATGTGAT CATCCCTTA ATGATATCTT AAGACCTCTA ATAGAACAAAT
12751 TTGGTCATGG ACTGTGGGGT TTTGCCCT CATTGTGTCA GCACTGAGCA
12801 TATTGTTGGC ATAGGAGGGGA TATTGTTGA ATGAATTGCT AGAGGTGGCC
12851 AAGAGATATG ATGTAAGTCA GGTTTTCCC TGCCCTTCCC CTTCCCTTC
12901 CCCACATCCT TCCTATAGCA GCCACCGTGG CTGCAGTTAC TGTAATGGC
12951 AAGACGGAAT CAGTCCCGA CATTGGGTTG TTTAGAAAA TTGCCTGCAA
13001 GTGTCAGGGT GATAAGTTAA AGCTTTGTCT TTTGCCCTCA GAGGAGCTAT
13051 CCCATAGTGA GTAGAAGCCA GAGAAGCTGA CCCCAGGAGT CTTCTTTCC
13101 AGCAGCAGGT CTTGAGCTGC ACTTCTCTGT AGCTACAATC CAGGCAGGAA
13151 CAAGCCCTAG GTACCTCCGG AGAGGAGGGC AAGAGAGGAA GAATGAGTTC
13201 AGCTACTCTA GCCACCAAAC TGATTATGAA TTGCCCTGAA ATCTGAAAAA
13251 TTCAATTCC AATCGTAAGT TTGTTTTGTT TCATTTGTT TTCTTAAATT
13301 GTATATTGTA AAGATGGCAT TAACTAAAGA TATATATTCA ATATAGAGTG
13351 GAAAAAAATGG AATACTTGCA TAGTATCTT TACTTATAGG TGATTTATGA
13401 TGGGGAGTGG GGTGGATAGG TTGGCAGTTC CCCCAGAAG TTGGAAATGA
13451 AGTTTGTCT CTGTGAGTTG AACTAATTAG ATCCACAAGT AATGAAAGCA
13501 GTATTGTGTT GTAGTTAAGA GCACACTCTA GAACCAGATT GCTTAGTTT
13551 AAATCCTGGT TCTGCCCTTT ATTATCTGTG TACTTTGGC AAGTTACTTG
13601 CCCTTTGTGT GCTTCATTTT TCTCATCTAG AAAATGGAGA GGCCAGGC
13651 AGTGGCTCAT GCCTATAATC CCAGCACTT GGGAGGCCGA GGCGGGCAGA
13701 TCACCTGAGG TGAGAAGTTC AAGACCAGCC TGGCCAACAT GGTGAAACCC
13751 TGTCTCTACA AAAATACAAA AATTAGCCAG GCATGATGGC GGGTGCCTGT
13801 AATCCCAGCT ACCCAGGAGC CTGAGGCGGG AGAAACACTT GAACCTGGAA
13851 GGCAGAGGTT GTAGTGAGCC AGGATTGCAC CACTGCACTC CAGCCTGGGT
13901 GACAAGAGCT AGACTCAGTC TAAAAAAAC TAAAAAAAC AACTGGAGA
13951 TACAGGCTGG GTGCAAGGCT TACACTTATA ATATCAGCAC TTTGGGAGGC
14001 CTAGGCGGGA GGATTGCTTG AACTCAGGAG TTTCAGATC AGTCTGGGT
14051 ACAGAGCAAG ACCTCATCCC CACAAAAAT CAAAAATTAA GCCAGGCATG
14101 GTGGCTCATG CCTGTGGTCC CAGCTACTCA GGAGGCTGAG GCGAGAGGAT
14151 TGCTTGAGCC CAGGAGGTTG AGGCTGAGT GAACCATGAC TGCAACACTA
14201 CATGCCAGCC TGGATGACAG AGCAAGACCC TATCTAAAAA TAAAAAA
14251 AAAGAAACGA GCCAGGCGCG TTTGCTCACG CCAGTAATCC CAGCACTTTG
14301 GGAGGCCAAG GCAGGTGGAT CACTTGAGGT CAGGAGATCG AGACTAGCCT
14351 GGCCAACATG GTGAAACCCC ATCTCAACTG AAAATACAAA AATTAGCCAG
14401 GCATGGTGGC ATGCTCCTGT AGTCCCAGCT ACTCACTTGG AGGCTGAGGC
14451 ACGAGAAATCG CTTGAACCCCA GGAGGCGGGAG GTTGCAGTGG GCCAACATCA
14501 TGTCACTGCA CTCCAGCCTG GGAGACAGAG CGAGACTCTG TCTCAATAAA
14551 TAAATAAACAA TAAATAAAAA TAAATAAAAA TAAATAAAAA TAAAAAAATA
14601 TGGAGGCCAG CAGGCACGGT GGCTCACCGA TGTAATCCCA GCACTTGGG
14651 AGGCCGAGGG GGGCGGATCA CAAGGTCAAGG AGATCGAGAC CATCCTGGCT
14701 AACACAGTGA AACCGCGTCT CTACTAAAAA TACACAAAT TAGCCAGGCA
14751 TGGTGGCAGG CACCTGTAGT CCCTGCTACT CAGGAGGCTG AGGCAGGAGA
14801 ATGGCGTGAA CCCGGGAGGC GGAGCTTGCA GTGAGCTGAG ATCGCGCCAC
14851 TGCAGTCAG CCTGGGCGAC AGAGCAAGAC TCTGTCTCAA TAAAAAA
14901 AAAATGGAG GTTGGGCGCG GTGGCTCGCG CCTGTAATCC CAGCACTTTG
14951 GGAGGTCGAG GCAGGCCGGAT CACCTGAGGT CAGGAGTTCC AGACCAGCCT

15001 GGCCAACATG GTGAAACCTT GTCTCTACTA AAATTACAAA AATTAGCCAG
15051 GCACGATGGC AGGCACCTGT AATCCCAGCT ACTTAGGAGA CTAAGGCAGG
15101 AGAATAGCTT GAACCTGGGA GATGGAGGTT GCAGTGTGCT GAGATCGCGC
15151 CACTGCCCTC CAGTAGAGTG AGATTCGTC TCAAAAAAAA AAAAAAAAGAA
15201 GAAATGGAGA TACAAACCTT CTACCTACCT CCTTACAACC TACCCCTCACA
15251 GTATTACTGT GAATAAAAGT GTGTGTAGCA CTGGGAACAC TATTACAGA
15301 GCACTCATGA ATGTTTGTTC TTTGTTATT GTTACTAGAG AGGCAAATGT
15351 CTGCCAGGGC TGAAATAATAT GTGTGAATTG GTGATTGTCG CACATATCTA
15401 AAGAAGTAGT TATTTTTTTC AATTAAAAC TAGTTAAAA ACCAATATAA
15451 GGCGGAGCGC AGTGGCTCAC ACCTGTAATC CCAGCACTTT GGGAGGCCGA
15501 GGTGGCAGA TCATTTGAGG TCAGGAGTTC GAGACTAGCC TGGCCAACAT
15551 GGTGAAACCC TGTCCTGCT AAAAAAAA AAAAAGTACA AAAATTAGCC
15601 AGGCATGATG GCAGGGCCCT GTAAATCCCAG CTACTGGGA GGCGGAGGCA
15651 GGAGAATTGC TTGAACCCAG GAGGTGGAGG TTGTTAGTGT CCGAGTTGT
15701 GCCACTGCAC TTCAGCCTGG GTGACAGAGG GAGACACTGT CTAAAAAAA
15751 AAAAAAAA ACCAAAACCA ATATAATAA TAAGTGGCCA GCAATGAAAC
15801 AGAAAGTGA AAGTTAGTGA AGCAAAACTA GTACTGTATT CAGATAAAGA
15851 TGCTGAATCT AGATTTGGTC ACCAGAAATAG GTTCTTTGT GGCAACCTGG
15901 GCTAGTTGG CTGACTCACC ACTGCCAGGA TGAAATTCTT TTCACTGGCT
15951 ACTCATTTC CTTTATTATA AGTCCATGCT CACAGAGCAA CCTTCTGATG
16001 CCTAATTCACTG CTTCTGGGA TACTTAATAA CAGGAAGGGT CTGGAAGTAG
16051 TACCTGTATA GGGGATATGA GTGTTCTGAT TTTAATAGTC AATTCTAAAG
16101 TGTACAGAGG GTTTGATAAA TGGTTAGGTC AGAACCATCA CAGAATGCT
16151 ACACCTCTT GGACATTAGG AAGGTCAAAA ACCTGAAAGG CCAAAAGCTA
16201 GGCCTAGATT AGGGTCATTC ACCAAGAAAA CATAGCCTT GAAGAGTTCT
16251 CTGGGTGGTC CACCAGCTAA CCTTCCTTGT ATCACACCTC CTTCTCGTT
16301 GCTTCTTAA GCATTGACCT GTAAATGGTA TGGAAATTGT TGCTCACCTA
16351 ACTCCTTCCTT TTACAGAGG AAGAAGTTGA AGCCCAGAGA GATTTAATGG
16401 CTTGCCTAAG ATCACACGCA GATTTCTGT TAACCAGGGT GATTTTCAAG
16451 GTGTTCCCTG CCAGACGAGG GCTTTTTC TTGAATTGCC TAGAGATTTC
16501 TTGAGATATC CGAAGCATT TTCCCAGTGC AGCCTGGAGA AGGATGTCCC
16551 TGTCAACACA GCATTTGTTA CTCATGTTA GACATTAAT TTTCTAATT
16601 GTATCATGGA GCAACAGTGG ATGATTATCT ATAAGGGGTT GCAATTCCAT
16651 GCTTATGTGC TTACAGCCCA TATAGACAAA TATCAGCTGT TAAAATGACA
16701 AGGCAGTAGA GATGTGGCCC CAGGACAAAG GCATACTCTG CTGTTAGTGA
16751 ACACTAGTTG GCCAGCAAAT TTCACATGGG CATATACAG GCCAACTGTA
16801 GACTTTAGGC ATTATACCC ATTCAAGAGAG CCAAACCTGGC AACTAAAGAT
16851 CAGCATTCTC TTTGGCATT CAGCTTGC TTCTGTTAAA AATCACTGCT
16901 TGCTTAAATA CCTCTGATAG CTCTTCAGT CCTGTTAGGCA ACTCTTCTGC
16951 CTAGCAGACT TGGTCTTTAG TGCTCTGCC CTACTCTCTT CCACCATTCT
17001 GGCCTCTGT CTAATTGCTG CCCATATGTG CCATGCACTA GAGCTTACAG
17051 ACCTGCTCAG CGTTATATGA GCATACCCATA CTCTTATGC CTCAGTGCAT
17101 TTGCACATGT TGTTCTTCA GGCCAGAATG CCTGTTACTG CCTGGCAATC
17151 AGCCTATTAG AGTCTGCCA TACCATCCCA TCTTCTGTGG AGGAGCCCC
17201 CGCCAAATCC ACCCATACCT CTCCCCACCA ATCAGAGACT TCTTCTCT
17251 TTGTTATTCT CTTCGTTATT CTCTTCATAC CTCAGTTATA TCCATTTCAG
17301 TATTTGTTTA CACATCTAGC ATCACTCTTA GAGTGTGAAA TTCTCCAAGT
17351 GTGGAGCCGT ATCTAGTTG TCTTTGTATC CCAGAGCTTA GCAAAGTGCC
17401 TAGAATGTAG TGGGTGCTCA GAGTGTGTTGC TGGGTGAATG ATGTATTGT
17451 TGAACGACTC TTTGGACACT TGAATAAAAGT CCATCCAGTA TGACCACTTA

17501 CCATCTCTTC GCTCTACAAT ATTCTTTAG GCAAGAGCTT ATCTTTTGTAG
17551 GTGATAAGAT AAGCTAAAC TTATGTAGAC TAAGACCTCA GTCTGTAAT
17601 GTCATCCCTA AGTCTTAAAC CATAAAACC AGGGCCTCAA GGAATGGCAT
17651 GCCTTCTGCA ACTGTAGCAA CCTGCTGTGC TTATTTGCC GTGTTTTCA
17701 TTTTCCCCC AAAAGCTAGA GTCCCTTCTC CCATGGGCAG TGCTGGAAGT
17751 GTGCTAACAA ATTCTTCTC CATACTGCTT ACGATTACAA AAAAACCC
17801 CAGCATCTCA TGCCAGACTT GAGTTAAGGT TGTTTCTTT TGTTGTCAG
17851 CTGTATTCTG GTCATGACTT CCTGATGATG CCCTATAGAG ATTTTGCTGA
17901 GATCAGAGGG TGCTCCACTG CCATCAGTAG CACTGACTCT TGCAAGAGCA
17951 CCGTTCTGA AGTTGGCTAA TGTCATCCCT CACGTTGTT TGTTGAAAT
18001 TTGTTTGTG TCCAGAGATA GCACTTCAT GGAATGACGC TATTTCTAG
18051 AATCACTTTT TTTTTTTTG TGAGTTGGAG TCTCGCTGTG TCGCCAGGCT
18101 GGAGTGCAGT GGCACAATCT CAGCTCACTG CAATCTCCAC CTTCCGGGTT
18151 CAAGTGATT CCCTGCCTCA GCCTCCCGAG GAGCTGTTAC TACAGGCCA
18201 CACCCCCACT CCTGGCTAAT TTATGTGTT TTAGTAGAGA CGGGGTTTCA
18251 CCGTGTGGC CAGGATGGTC TCGATCTCT GACTTTGTGA TCTGCCTGCT
18301 TCAGCCTCCC AAAGTGCTGG GATTACAGGT GTGAGTCACC GCGCCTGGCC
18351 TAGAATCACC TTTTATACC ATAACGTGAG CACCACTGCC GCGTCACCAA
18401 GGAAAGAGAG AGGCAGCTAC TGTTGGGTTA CAAATGGGTA AGAGTGGCAC
18451 CAGGAAGGTG AAAGTCTCTA CTTAGCCAAG GCTTAACAAA ATGTCATCA
18501 CAAACATTT ATTATTAAG CTACGTTCA GATAAGAAGA TGAACAAGCT
18551 ATCTGTACAT TCATTTCTC GTTGTAAAC AGGTATGAT AGTGTATCTAT
18601 CCTGCCTGCC TCTGAGGGTT ATTGTGAGAA TAAAATGAAA TCAAGTGGAA
18651 AAGCACTTAG GAAAAAGAAA AGCATTGGTT TTCAATTGTT AGTGTGGATC
18701 AGAAAACACTG GGGCTTGTG AAAATGCAGA TTCTTAGCCC CAGTCTCAGC
18751 GATTCTGATT CTGTATATCT GAAGTGGGAC TCAGGAATCT TGATTTTCAA
18801 CAAGCTGACC AGAGGGTCCA ATGCTGCTAT TCCTTAGTT ACACTTTCAG
18851 AAATATTACT GTAAATCAA TGCAAGAAT AAAATAGTTA TTTGAGGCAG
18901 TTTTAGTATG TTGGACCTGG AGTCCAAAGA CTTGGTCAA ACTCCAGCTT
18951 TGTCAGTCC TAGACCTGTG ACCTTAAACA GCAACCTCT CTGTGAACCT
19001 TAGTCCCTC AGGAACGGCT CTGGTCACCT CCTGCTGTAC TCCATTGATG
19051 ACTCACCA TAAGGCTCCC TGGGAGTCCC CCAAACCTT GCTCTCTTAA
19101 CTCCTTTTAC AGCCTCTAC ATCTCTGCA GGTGCTGTCT TCTCCTCCTT
19151 TTTCCAGGCC CTGCTCTGAC ACAGCATTCA TTCTCTCTG GGAAGGGTT
19201 CTTCAATGTG TCTCCAAGCA CATCACACCC AGGAAGGACC CTGTGGCCAT
19251 ATCTGTCTAT CACCAAGATCA AACTACGTGA AGGCAGGCAC TAGGTACTGT
19301 CAGTGCCCAG CATAGGCCTG GCCCATACCA GGTGTCACA GATGCCTAGT
19351 AAAGAAACCT ATGATTCAAGG ACCCCCATGA TGAGCAACTA TAGCACTAGA
19401 ACAGTGATAA TAACTAATGT TTATAATGCA TCTTCAGTTT ACAGAGGGCT
19451 TTTGTACTCA TCATCTAGTT TAGTTCTGC AACACCTCT TGAGGAATAT
19501 AGCACAAGCA GGACAAGGGGA AGCCCAGAGA TGTTAAATAA TTTATCCAAG
19551 TTTATGCTGC TGGGAAGGGC AGCACTGAAA TTAAAAGAAA AGTTTTCTGA
19601 GCTCAAATCC CATGCCCTT CCTCAATGT AGCTCTAGCA AGGTATTGAG
19651 GAATCCTGCC TCTACAGTTC AGAGCCTCAA ATTGCTGGGT ATGTTGAGTT
19701 CTTGTATCTG ATTTTCTAG ATTTCTGCA CACATTCTTA CTGTCTGGAT
19751 ATCAGGAAAG AGTTTATCAA ATGCCTGTGG AAATCCAAGA TAAGGTCTCA
19801 TGATGAGTAA CCCAGTGAAA ACATGAAGTC AAGTCTAACT AGTCACTACT
19851 ATTTCACTAC TGCTGACTCC TGATGATCAG CTCCCTTCT AAGTGCTTAC
19901 TGTCCACTTA TTCCATCATC TGCCCTAGAAT TTATGTGAAG GAATCAAAGC
19951 AAAAGGATCA TAAGGCTTCC TTTTCCAGT ATGTTTTCC TCCTTTTGA

20001 AAACTGGGCC AGTTAGCTAT CTCCATTTC ATTTCATGAA TACATCCCCA
20051 GCGCCTGGTA TATAGTAGAT ATGGAACATT ACACCTTGGA GATATTGCAC
20101 CCATTCTCCA GTTCTCCAA AGTTACTAAC AATGGTTCCA TCACTGTGCC
20151 AACATATTTC CTTTTTCAA TATATTGGGA AATAATTCTC CCAGTCTGAA
20201 AATCTGAACA CATTCTATGT GACTTGGTAT CCTCATATGT CTTGGGCTTC
20251 CAATTCTCCA TTCTAGTT CAAGTTCATG AACTGTAAAA CAAAGGATTA
20301 GACTAAATCT CAAAGTTCT ATCCAGATGC CAAATTCTT TCTCTTTCCA
20351 TGATACCTAA GATAGATGCC AAATATTGTC TTTTACCTGG TGTTTGTGAA
20401 CATGACATCA CATTACAGGA GTAGCAGATA CTAAACTCTC ACTCTGTAAA
20451 AACTGACTG AGTTCATGA GCCAGATACT GAAGTGGACT TGTTCACATA
20501 TGTTCTCATT TAATGCTCAT AACCTGTGA AGCTGGGAAT TGCTGGGACA
20551 TTTTATTAT TTATTTATTG AGACGGAGTC TGGCTCTGTC ACCTAGGCTG
20601 GTGTGCAATG GCATGATCTT GGCTCACCGC AACCTCCGCC TCCCGGGTTC
20651 AAGCGATTCT CTTGCCTCAG CCTCCGCAGT AGCTGGGATT ACGGGGCACA
20701 CACCACCA CCAAGCTAAT TTTGTATTT TAGCAGAGAT GGAGTTCTC
20751 CATGTTGGCC AGGTTGGTCA CGAACACTT ACCTCAAGTG ATCTGCCTGC
20801 CTCAGCCTCC CAAAGTGCTG GGATTACAGG CATGAGCCAC CATGCCTGCC
20851 CGGGACCCCTT GTTTAGAAG GATGACTGCT GCTATAATGT AGAAAGTGT
20901 TTGGAAGAGG GGAGGAGTGG GGCACGAAAG ATGGTTAGTA GATGGGGGTG
20951 GTAATGCTTA CCTTTCAGTA TTTGGAGGCT TCAGGAGTCCT CAAAAATTCT
21001 CTTCTTGTAT TGGAGTCCTC CCAGCCAATA GAGGGCTTC CACAAACAGT
21051 TTCTTGGGTT TTGAATTGTT TGACCAAGAGC TTTCTCCGA CAAAAGGTTG
21101 GGGTGATTCA TTCACCTTAC ACACCTTGCC TGAACATTCA CTTGGGGCTG
21151 CCGGTTATGA AGGCTATTGT TCTCCAGCCT GTCACAGACG CTTTGAAGAC
21201 CTGTGCCTCA GCTGGTTCTA AGGAGTCAGT TTGTTAGCT CGTGCAGG
21251 TTTCCAACCTT ATGAAATGTG CTGGAGATTA ACACCTCTCC TGCCATTTC
21301 TCCCTACTAT AATTGCCAGT CAAAGGATTG CTGAGTTGC CTCTGGCAGC
21351 CATAACTGAT GAATGTTCTG CCAGCTGCTC TGAGGACCTA GAAGAGCAGT
21401 TTTCTATCCA GGACCAGTTT CCAAGGGTGG GAGGGTGAAA TATATCCTCC
21451 AGTGTGACAT TTCATCTCCC AGTGTGGGT GGCTTGGGCC CTTTGAAGTT
21501 GGCTCTGAGG AACACACAC TTGGGTCTGA GCAGCCAGCA GCTTATCACA
21551 TCTGGTGATC AATCCTCAA AGGTTCTCC TGAAGTCTGA ATTGGGAG
21601 GTCAAATGGA TTCCACCTGG GAGGGCTTC TGCTTCAACT CAGGACATGG
21651 GGAGAAGGCT GTTCCCTTCC CAGGGGGAGG CAGTTTCAT GGCATTGAGA
21701 TGTCCCTCTCA CTTATTCCCC ACCCACCCAC CAAGTCTTT GTAAGAGGAG
21751 TAGGGGGAGA GGAGAGCGCC TGCAGCCTCC TGCTCACATT CCTAGACACC
21801 GACTCACTGA GCCCCTGCC GCTGGAAACAG CAGAGCTGTG TGAATGTCA
21851 AGAGGAGTTA TGCTCATAGG CTCCTGGCC TCAGTCTCTT TGTGGCTTGC
21901 ATATTCTTCC ATTAGTACTG TGTTCATCAC ATGAAATCA GAGGGTACAA
21951 TTAAAAGATA ATTGCTAGT CCCAGACTTA ATTTGGGGCC CCCTTCTTGC
22001 CTGATTGAAT TACAGGGAA CATAATAGAT TTTTGGTGTAA AAATAGTTGT
22051 CTGTGTGGCT GGGGAGAAAGA TTGCTCCAG CTCTCCAGCT GGGCAGCCCT
22101 TTCACTGATCC CGTATGTTAT TTCCCCACTT CCAGCCCACC TCACCTCTC
22151 TGTGGCCCTT GTGTGTCCCC TCGGCTAGGA TCCTGACCTC CTGCTCAAGA
22201 GTTTAAACTC AACTTGAGAC CCAAGGAAAA TAGAGAGCCC TCTGCAACCT
22251 CATAGGGGTG AAAATGTTG ATGCTGGGAG CTATTAGAG ACTAACCAA
22301 GGCCCAGACA GAGAGAGTGA CTTGCTAAAG GCCACATAGC TAGCCCACAG
22351 TAGTTGTAAAC AATAGTCTTA ATGATATTAA TGGCTAACAT TTATCAACCT
22401 TTAATGTGTC CCAGACTTTG TGCCAAGGGC TTACATGCAG TGCATTGTG
22451 CATTCAAACC CAGACAGTCT GGCTCTGGC CCAGGCTGAG CTTGGTATA

FIG.3-9

22501 GCATGGTAGA ACGTTGCTA TAATGCTAG TCTGGGTTCA AATCCTGGCT
22551 TCACTTCTCA CATTACAGC TGAGTGACCT CAGGCAAGTG ATTTAACCTC
22601 CCTGTACCTC AGTTGCTTTA TCTGTAAAGA GAAAAATCAC AGCACTGTGG
22651 AATAGTGGGG GTTAAAATTC ATTACATACAA GTAGTGCTGC AAGCAATGTT
22701 TAATACAGGG TGAGCACCTG TTCAGTGCTT CCTTCTCTG GCTGCCTCTG
22751 GGGCTAGAGT GTGGTGTCTT CGTGGTATAG ATAGATAGAT ATGGCTGAGC
22801 TCTGCACAAA CACCAAGAGC TGTTCTTCAC TATTAGAGGT AGTAAACAGA
22851 GTGGTTGAGC TCTGTGGTT TAGAACAGAG GCCGGCAAGC TATGGCCCAT
22901 TGCCTATTT AATACGGCCT GTGATTGATT GATTTTTTT TTCTTTTTGA
22951 GACAGAGTTT CACTCTTGTG GCCCAGGCTG GAATGCAATG GCACGAACTC
23001 AGCTCACCGC AACCTCTGCC TCCTGGGTT AAGCGATTCT CCTGTCTAG
23051 CCTCTCGAGT AGCTGGGATT ACAGGCATGT GCCACACGC CTGGCTAATT
23101 TTTGTATTT TAGTAGAGAC AGGGTTTCTC CATGTTGGTC AGGCTAGTCT
23151 CGAACTTCCA ACCTCAGGTG ATCTGCCGC CTCAGCCTTC CAAAGTGCTG
23201 GGATTACAGG CGTGAGCCAC CATGACTGGC CTGATTGACT GATTTTTTA
23251 GTAGAGATAG GGTCTTGGTT TGTTACCCAG GCTGGTCTCA AACTCTGGC
23301 TTCAAGCAGT CCTCCCTCCT TGGCCTCTCG AATGCTGGGA TTATAGGCAT
23351 GAGCCACTAT GCCTGGCTA TATGACCTGT GATTTTAAT GGTTAGGGGA
23401 AAAAAAGCAA AAGAATGCTT TGTGACATGT GGAAATTACA TGAAACTCAA
23451 ATATCAGTGT CCCAGCCTGG GCAACAAAGT GAGACCCGT CTCTACAAAA
23501 AATAAAAAAA ATAAGCCAG GGCGGGCGC AGTGGCTCAC ACCTATAATC
23551 TCAGCACTTT GGGAGGCCGA GGCAAGTGGA TCACCTGAGG TCAGGAGTTC
23601 AAGACCAGCC TGACCAATAT GGTGAAACCC TGTCTGTACT AAAAAACACAA
23651 AAATTAGCCG AGCATGGTGG CATGCGCCTG TAGTCCCAGC TACTTGGGAG
23701 GCTGAGACAA GAGAATTGCT TGAACCTGGG AGGCGGAGGT TGCACTGAGC
23751 CAAGATCGCG ACACTACACT GCAGCCTGGG CAACAGAGCG AGACTCCGAC
23801 ACACGCACGC ACGCACACAC ACACACACAC ACACACACAC ACGCTGGTA
23851 TGGTGGCCAG CACGTGTGGT CCCAGGATGC ACTGGAGGCT TAGGTAGGAG
23901 GATCACTTGA GCTTAGGTGG TTGAGACTAC AATGAACCAT GTTTATACCA
23951 CTGCACTTTA GCCAGGGCAA CAGTGTGAGA CTGAATCTCA AAAGAAAAAA
24001 AAAAAAAAGA AAAAAATCTT TCCATAAGTA AATATCTGTT GGAACATAGC
24051 CATGTCCTT AGTTTATGTT TTATATATGG CTGCTTTGC CCTATAATGA
24101 CACAATTGAG TGGCCACGAC AGTCTGTATG GCCTGCAGAG CCTAAGATAT
24151 TTGCTCTCTG GCCCTTACA GAAAAAGTGC CTTGACCTGT GCTCTAGAGC
24201 CATATGTACC AGGTTTGGAA CTCAGCCTCA CAGCTGGGTG TGATGGCAGC
24251 CATCTGTAGT CCCAGCTACT CTGGAGGCTG AGGTGAGAGG ATCACTTGAG
24301 TCCAGAAGGT CGAGGTCAAG ATTGTAGTGA GCCATGATGG CATCACCGCA
24351 CTCCAGCCTG AGTGACAGAG AGAGACCTG ACTCAAAAAA AAAAAACAA
24401 AAAAAAAACCA CACCCCTCACC ACTTATCAGC TATTTGTCTT GAGAATAGT
24451 ACATAACCCC TCAGAACCTA TTTCCTAAC TGTAAATGA GGCTGATGAC
24501 GTTTCTCCT TTTACTGGCA ATTTAAACAT GATGGATAAT AAATGCTAAG
24551 CACTAACAC AGGGCCTAGA AGATATTAAC TGCTCAATAA ATGGTAGCTT
24601 CTTAACAGTA TCAAACCCA TGTGCTTTA TCACATGCAT TGTGTCCT
24651 GTGTCCAGTT GGTGGAATGG GAAAAGGCTC CCTTGTAAACC CCATCTACCA
24701 TCTTATCAG ACTTTCTGC CATGGTTCAC AGTAAGAGAT AGAAGCTGCA
24751 CGGTGACTTC TGCTCTTTA CAATGGTGAG CGGTGTGTGC CTGGTAAGGG
24801 AGAGCTGATG TCACTGCCCA AAATCCAGTA GTGAGATCTG AGTGTCTGG
24851 TTTCCCTCAG CAGCCTTGCT TTTTCTTTA CAATCCTGCA GGCAGGGAGA
24901 CAAGGGCTTT CTACATGGTA GGCTCTGGTT TGGTCATCGT CACAACCTGGG
24951 GGCTGTTAG GTGGGCTCCC ATTCCAGATA CCTAGGCTTA TCAATCCCTT

FIG.3-10

25001 TTGGCACCCC AGGCCTTTT CTCCCTCATG CCCCCATTTT CAGTTTGAAA
25051 AGCATGGTTA TCACAGGACA AGTAGAAGAA GCTCCACTGT CCACTGAGGC
25101 CAATGGATGG TGTTCTGCAT GTGAACACTC AGTGAATAGT GAGTGAATGA
25151 GAGTAACCTG GGCTCCATCC TATTTGCAGA GAGCTTGGA AAAGATTTT
25201 CTCCCTAAAG AGCCAGAATG AAGCCTGGTA GTGGGAGAGC TCCAGCTCTA
25251 GAGTCACATG AGCCTACATT TAAATTCCAG CCCTGCCACT GACTCCCTT
25301 TTGACCTTGA GTGAGTTACC TAATCTCTCT GTACCTCACT TTTCTTGCT
25351 GTAGAGTGGG AATAATTCCCT GTCTCAGAGA AATAAAAGAG TGCAATATAGT
25401 GTTGCCACA TGGAGACACA TCAGGGTAG GTTAATACTC TGCGCCTTGT
25451 TTCCCTTATT GCAACACAGC CCTGCCCTGG AGTGGAAAGTG GCACCTCCCA
25501 TTGGTCAGCT CTTGAGGCTG TCCCCAGGAC AGGCAGAGGG AGGGAATGAA
25551 TGGGAGCCCT AGTGCCAGGA CAGAACAGAT GGCAGCTCAG AGCTAGGATG
25601 GCTCTCTGGA CCTGTCTCTC CTACCAGAGG TCCCCCGTC TGTTGTGGCT
25651 CTTCCTGGAC CTGGCATCCT CTGCTTTTTT TTTTTTCCA CCTCCAAGCA
25701 GAATTACTGT CCTGTAGGCA GCTCCTCTGC TTGAGGACAT CTGGGGGCCAG
25751 ATATGTTCAC ACTCTATCCT GCCTTGCCCT TCCCTGAGCT CAGGATGGAC
25801 GCTCAATTGG TCCCAGTTAT TGTCTGCAGC GCCTGCCCTGC AGCCTCGATC
25851 CAGCCCAGCT CCACCCCTTG CCTGCAAGGT CTGTTTCCA ACAGCTGCTC
25901 CAACCACACA CCTCGGTTCT GCGGGAGCCC CTCCCTTCC TCCCTCCCTC
25951 CCTCATTCAG GGGTGGGACT GAAGAAGAAG GCTAACTTGA CAGCAGCGCT
26001 TCTTTCTTAG CTAGTCACCG GCCCCCTGCTC AAGAAATGCCA GTGTGTGTGT
26051 AGCCTCCACA GAGAGGTGCTG TTTCTCGGAG TCCAGAGGGG CCGCCTGAGC
26101 TTCTGAGAAC TAGGGAGGAG CCATCCCAGC CATGAGCCCC TGTTGGAATC
26151 TGCTGGGGC CAAGTGGCCT GGAGTCTCAG GCCTCCCGCA GCTGCTCCGG
26201 AGGGAGAGGT GAGCTCAGGG CAGCCTGCCCT GCAGCCAGAG GTGCCGGGAG
26251 CCCCCGGGCT GTCATGGTGG CCATCTACAG CCGGCCTGAG GCAGTCACAG
26301 ACGGATTTCG AGCTGAGCCT GTCTATCTGG TGTGGGAAGA AGATGGGGAG
26351 TTACTTGTCA GTCCCGGCTT ACTTCACCTC CAGAGACCTG TTTCGGTGAG
26401 TTGGTCTCCG AGTCCCCCTC TCCATCTCTC CTGGCCCCCTG GTCTGAGAG
26451 GAGGGTGGTC TCCCTAAATC TCCCTCTCAC TTAGTCTTTT ACCATCGGTT
26501 CTGCCGGGCA GAAGCCAGCG GAGGTTATAC CCAAGGAGAA TCGGCCCTTGT
26551 GAGGTACCCC CATTATGTCC TGGAAAGTGGT GAGGGGAGGG ATATAACCCAG
26601 AAGGAACCTTC TTAGGGAGCT CCAGCTCCCC TTCTATCCA GACAAACCTG
26651 AAGGAGCCTC CAAAGATGC CACTGACCTG CCCATTGTAG ATGTTACTGC
26701 TTCCGGGGGG AATAGCCCAA ATAGAGTGCT GTTTCAGCT CTCACATGTC
26751 TTACCTGCGG GCCATGCTGC CTGCCAGGA ATTGTCCA ACAAGCAGGA
26801 TGGGCAGGTT TTGCAAACACT GTGGAAACTG GCAAGTCTG GGTGTGGTA
26851 GCCTGGTACA CAGTAGGCAC CTTATAAAACG TTGTTCTCT TAATGGCAGG
26901 CACATTGCC TCTGCCCTTG AAGGGCTTCT GAGCTCCAG GTGAATGTAG
26951 TTGCTGGGGGA AAGACCTGGG CGAGTGCTTC TAAGACTGGA GCAATGGGCT
27001 TTAGAGTGTGTT CCTGAGCTGC TGGGCCAGCC CCCACACCTC CTCAGTCCCT
27051 AGGCCTAAAGT ACCTCCACGA GCCTCTCTCT GTGGGGCTTC TCAGAGGGAG
27101 ATGTGGAAAC TCTACCTCTA ACCTGGCTTTT CTTCGCTCAT TGCCCCACTC
27151 CACCTCCCAT AGAAACTCCC CAGGGGGTTT CTGGCCCTCT GGGTCCCTTC
27201 TGAATGGAGC CATTCCAGGC TAGGGTGGGG TTTGTTTCA TTCTTTGGGA
27251 GCAGCCTGTT GTTCCAAAAA GGCTGCCCTCC CCCTCACCAG TGGTCTGGT
27301 CGACTTTCC CTTCTGGCTT CTCTAAGCTA GGTCCAGTGC CCAGATCTTG
27351 CTGCCGGGAT ACTAGTCAGG TGGCCAGGCC CTGGGCAGAA AAGCAGTGT
27401 CCATGTGGTT TTGTGGAATG ACCGGACCCCT GGTAGATTGC TGGGAAGTGT
27451 CTGGACAGGG GGAAGGGGG AAGGAAGTGG TCCTCAATGC TGACTCTACC

FIG.3-11

27501 AAGGCCCTG CTAGACACTT TATCCTTAA TCTCTAACCA GCCTAAAGAG
27551 ATTATATATC CCCATTTAC AGATGAGGCA ACCAGTTCA ACAGAGTTAA
27601 CATATGGAGC CTCACTGGGC AGCTTTTCT GTCTTCTGA CTTTCTCTCA
27651 TCCTTCAGGG GGCTGCAGGT TTGTTTCTT CTCTAGTGG AGAGGAAATT
27701 CTCAGGTTTGT TTTTCTCTC CTAGCAGAGA GTAAAAAAAG GGATAGTTG
27751 CCTGACTTGTG TGAAGGTGTG GCTGAGATTG TTTCTAAAG AGCCAATGG
27801 AATTGATCTT GAGTTTAGGA GAAAGCTTT ACATGTGGAA TTAAGATGCC
27851 AAGTGTGAA GTAGCCACAT TTCAAGGTCTT CATTAAATTTC TCTTAATCCT
27901 GGGAAAGGCAG CTTAGGAGAA GGGTTGTTCC TTTAGGAGCC AGGAACATA
27951 CCCCTTTAC CTTGGAGAG GCAGGGAAAGC CAGGGAGGAC ACAACCTTC
28001 AGGAAGAGGA GAAGCTAGAG CAGATAGTGA ACTCTAACCC TGAACCTTTA
28051 AGGGCCAGAC CACTAATGCC ACCCAAGTCC ACCTGCCGTT TGTCTTGTTC
28101 TGTCCCAGGC TTTCTGGAGA ACCTGATCTT CTGCCCCCTA CCCCCAAGCT
28151 CCGTTTGGCC AGCTAGAGTC TGGGGGGTAC TGACTGACTT TCGTAGACAT
28201 TCTTCCCTTC CCCAAATAAG AGGCCACATT CCTGAAGTCA CTTCTGAAGA
28251 GATAGCTGCC ACACAGGGCT CTTTCCCCC AGGGAGGGAC CACCCAGACC
28301 CTCTGCTCTC CCAGGTATCC GTTACACAT CACTACCTGG TCAGAAAGCT
28351 GTTTCTGCCA TTAGCCCTC CCTCTTTTAT TATAGGATAT CCTCAAGGGC
28401 TCCTCTTGG GCCTCAGTTT CATCCTTGGC AGAAAGTAGA AGCTAGACTT
28451 CTTGGGCTCC TGAACAGGGT CTTGCTGGA TTCTGTGAAA CAAATTAAGT
28501 TCTTGACCCCT AGGCCTCTGG GGGAGTACAA AGTCTATGGG AGTTCTGGGG
28551 CTGTGGTTGC AAGGAAAGTG ACGCAACCCAG ATTCCATGGG GACATGATCA
28601 GGCCTGACAT GTGAGGGAGG AAGAGGGAGC AAGGGAATGA AGAATACAAC
28651 TTCTGTGTCC CATAACCCCC TGCCCTGACAG GCCATACATA CTCAGCAGAG
28701 AATGCACTGT CTTCTTACCC ACACTAGCGT GAGGAGTGAG CTGCAATTAC
28751 CACTGTGCTT CCAAGTAAGA AAATACCTCA AATTGGATT TACAAAAGAG
28801 GTAAATTAGG GAGTGGCTTT TGTCGGACAT CTTTAAAGCA TTTTTCTTTT
28851 TATAGAATTTC CACTTAATGT CCAATACTGA TTTAATGAGC TTGGGTTTAC
28901 ACATTATCTC TTGAAGAAAA CAAATGAACC TTTGTGTTCC AAAGCAATTCC
28951 ATGTTAAAG GGAAAAAATT ATGCATAACT CTGCCCAGCT TCACAGTAAC
29001 CTTGGCAGG TGCCCTTAGGT CCTCTGGGAC TCTTTCCCTT ATCTGAAAAA
29051 TGAAGGACTT GGATCAGGTG AATGGTTCCC AGCTCTGCAA CTTATGTGGC
29101 TCCTCAGAGG CACACAAGCT CTTTCCATT ATTTGCCAAA TAATGGAGGC
29151 CCTGTCTTAA ACTGCAGTAC AACTACACAA AATACTGAA ACTACAGTCT
29201 TCCTGGTTT TGGTTGGAAC TGAATCAGTG CACTCTAGCA ACACCTTATT
29251 CTTGCTGTTG GTAGGCTTCA TTATGTGTTT GGTAAATTTC TTAAAACAAC
29301 AATAACATAT TCCATAATAA TTACAGCTTA ATTGGCAGAC TGTTTCAGTC
29351 TATAGGATCT GCAGGAAGGA GGAGTAATAA AGGGATTTTT GACTGAGCTC
29401 TTATGGAACA GAGTCTCTC AGGCCCTGT CATATCTGCC CTTCTGGGCC
29451 CTGGGGAAAA GTGGGCATCC CCAGTTGTTG TGCTCTCCAG GTGCCCTCAG
29501 GCTGTGGTGG AGGGAGCTTC CCATTCTCTC CTTCAGCCCA CTCATTCTCAG
29551 AGGCTAGGGG CTGAAAGAAG CTTCTCTACA ACTGGCTGTT CACTGGGAGG
29601 TTAAGGGATG ACCATCCAGC CAGGCCCTCC TCAGGACATG GGAGGGCTTA
29651 TGCTTTAACCA TGTGTAATC CACTGCAATA ATGACTGGTT CTTTACCCCC
29701 ATAAGGTTGA GAATTACCT GTAAACATT TTGCTGTGAG AATTGGATG
29751 TAAGTGGGG CTGGGCCTCT ATCTTATCTC ACTGGCTTC TCTCAGCACA
29801 GCACCTTGCC TGCTTGTCT TACACATCTC AGATGCACAG TAACTATTT
29851 CTAATTATTA GAAATCTATT AGAATCAATT GATTTCAGCT GGGCTTGGTG
29901 GCTCCTTCCT GTAATCCCAG CACTTGGGA GGCTAAGGCT GGAGGATCAC
29951 CTGAGTCCAG GAGTTTAAGA CCAGCCTGGG CAACATAGGG AGACCCCTGTC

FIG.3-12

30001 TCTACAAAAA ATAAAAAATT AGCCAGGCAT GGTGGTGTGC ACCTGTAGTC
30051 CCAGCTACTC AGGAGGCTGA GGCAGGAGGA TCTCTTGAGC CTGGGAGGTC
30101 AGACTACAGT GAGCAATGAT TGTGCCACTG CACTCCAGCC TGGGTGACAG
30151 AGTAAGACTC TGTCTCTAA AAAAAAAA AAAAAGTTG ATTTCTATTT
30201 GGATAGATAA ATAATTCTATT TTAGGACCTT TCTTTTCAC TTACAGAAAT
30251 CTGTTTCATT CTGGGCTGAG AAGCAGGTCC ATATTGCTAG GCATAGGAGA
30301 AAAAGGGGTC TGTCCTGCATT TGCCCTTGGT GGTCTCAAAT TGGGGAGGGA
30351 AAGAAATGAA CACTTACTGG CTACCTTCTG TGAGCCAGGC ATCATGCAAG
30401 ACATCTGTAC ATAATTAAAT TCTCATAACC CCATAAGATA TTATTAGCAA
30451 TGTACAAGTG AGGAAACTGA GGCTCAGAGT CATGAAGTAA CTGGCCTTGG
30501 GTGACACAGA TGGTAATGG CAGAGAAGGA ATATGGATCC AGGTCTTGAA
30551 AGAGAAAATC TCAACTGATT ATCTTTTTA AAAAACTCAT ATGTTCTCTG
30601 CTGACTCAAAGGCTCTGT GTGGATCTGG GTTGACCCAC TGAACGTGACC
30651 ATCAGGGTTC CATGCACTTT GTATCTGCC AAGCCCTCAG AACCCTCAG
30701 TAATGTTTTG GAAGATGAGT TTTGGAGGTT GTCTTAGGC ATAGCCTCAG
30751 CGTATGTAGG CCTCTAGGTG ATCTCCCTA ACCTGAGGAT TTCAGCTCAA
30801 TTCACTCTGG CTCTCAGGA CAGTGGGATG ACTGGTTAG ACCTCAGCTT
30851 TACCACCTCC CAGCTGGGTA CTCTTCTACC TACAGCCAGG GCAGATTTG
30901 ACTTTCACTT GAAACTTCCA AAAATTGAAA GGTAGAAAAA CAGCCTTGGC
30951 TTTGGGAAGA ACGTATGATG TCCATGGCCT CTAAGCATCT GAGGTGGGAC
31001 ATGTTCGAGT AGCACCTTAC AGTTCCAAAG TGTGTTCTGG GTTCTTTGTT
31051 TAAAAGAACAGAGACTGCTG GGGATTGAA CACTGTGAAG TATATGAAGG
31101 AGGAGAAATTG TGCTATTAA CATTCACTGAC TTGGGCTAAA GGAGAAGCAG
31151 CACGAAGTGT TAACACTCAA AGGGTCTTGA GCTGTCAAGGG CTCCAGCTTC
31201 CTTATTTCA CAGGTGAGAA TCCTGAGGCT CAGCTGTTGA GATGTGCTGT
31251 CTCACTCCGG TGACATAGTA CAGTGGATGT GGCTTGCAG CCAAGCACAC
31301 ATAGCTTCAC ATTCCAGCTC CATCAATTAT GTATTGGCA GCTTTGCAGA
31351 ATGATTTGAC TTAACTCTG CTTTCAGTC TTCTGTAAAA CAGGGATAAT
31401 CCTGCTACCG TAGGGTTGTC AGGATTAGAG ATAATATAAA TAAGGTACCT
31451 CATATAGGAC CTGGATTATG GCTGGCATTTC AATAAATAGT AGCTGTTAAT
31501 TGATAGCTAA GCTAGAACCTC TGAAGTCTAC CATGGCAACT TCTTAAGTGG
31551 TCTGAGAACCCAGTTGTCTG CTGTGGCAAA ACACAGCTTA GGGATCCATA
31601 CCCAGCCCTC CTGTCACTG TTACACCTTC AGTTCTTCAG AGACATGTGT
31651 GGCAGTGAATGTTGCCACAT AGCTGGCTGT GCCCTTAAA GGCATTCCCT
31701 GACACAGATA TGTGGACTGG TGACGTTGCT CTCCAGCCAG GTGTTCTTCC
31751 CAGCAGGCTG GCCTGGCTGT CTCCCTGCATG CCTGTACTTG TTTGTCTCCC
31801 TGCTCCCTCT CCTGGGCTG GCCAGAGCTA CTTGCAGCAA ACAAAAGCAG
31851 GATATTGGCA ATGAAAGGA GGGTGTGTT TC TGTTGCTCCC ATGCCCTGCG
31901 GCGCACATAC CATTGCAAGG GCGTAACAGA GCCCAGGCCT GCATTTGGGT
31951 GCAAATAAGT CTGCACACAG AAGAAAAGAA GGACCTGGTG ACCAGGAGCC
32001 ATGGAACCTCTGTGCTCCCC TACCTGGGCT ACTGGTTCTT GCCACTCCTA
32051 CCATTTCAAGTTGGAAATA TTTGTTAAGG CTTTGCTCTT CCAGGTCCCT
32101 TGCTTGGTGC TGAGTCTACC AAGAGTAAGT GGGATGCTGT TTTTGTCTC
32151 AGGGAGCTAA CAGTCTAGTG AAGAAGAAAG ATGGTTGCC AGGAACCTTCT
32201 AAGTCAGAAG GCAGGAGGCA AGAAGGAAGC CCCTGCTCCT ACTGCCAGCC
32251 CTCTGTTGGG CACCCCATAG TTCTTCAGAA CCACATTAA TCCTCACTGC
32301 AGGCCAGGCA TAGTGGCTCA CACCTGTAAT CGCAGCACTT CGGGAGGCCA
32351 AGGCAGGGCAG ATCACTTGAG GTCGGGAGTT CGAGACCAGC CTCACCAACA
32401 TGGGGAAACC CCGTCTCTAC TAAAAATAGA AAAATTAGCC GGGTGTGGTG
32451 GCATGCGCCA GTAATCCCAG CTACTCAGGA GGCTGAGGTG GGAAAATCAC

FIG.3-13

32501 TTGAACTCGG GAAGCAGAGG TTGCAGTGAG CCGAGATTGT GCCACTGCAC
32551 TCCAGCCTGG GCGATAAGAG CAAAATCCA TCTCAAAAAA AAAAAGAAAA
32601 AAGAAAAAAT CCTCACTGCT ACCTTGAAAG TAGGTGATGA CATTGCCATT
32651 TCACAAATGA GAAGTGAAGG GGCTAGCCC AGATCACTTA GGTGGTAAAT
32701 GGTGGTGCTA AGATTAGAAC CTCAGATCAT CTAGGGAAAA ACACAGATAT
32751 GCACAGAGTT AAGGGGACCC AGGGTATTGT TTGTCCTCTT GTTTCACAGG
32801 TGGGAAACA ACCCAGAGAG GGAAAGGGC TTGTCCAAGG CAATTAGCA
32851 CCCAAGAACT TGAACCCATA TCTCTCTCT CCTCATTAG AGCTCATCCC
32901 ACATGTATCT TATATTGAGA GGAGTGTGAG CCACATACCA AGAACAGTCT
32951 TCCCCCTGC CTCCAACCTC ACTGTGCACT TTTGAGACAC TTCACAGCCA
33001 TACTCTTCAT GCCATACCCA GCCCTTAAGA CCCTGAAGTT CCCCTTCAT
33051 AAGACAAGTA GGAAAAGCTA TAGGGTAAAA ATAGCCATCA GTGTTTGTG
33101 AGCACCCAGG AGGAATTGGG CACTCCAGAA AGATAAAAGGG ATTCTCAGGG
33151 ACTTGTCTCT CTAGACTTCC CTAGCTCAGC TGCTTCAACT CATTCTGCC
33201 CCTCTTCTCT ACCTCCCGCA GTGCTCAGAA GTAGTAGAAC TCACTGTGGC
33251 CTCTCACCTT GCATTGTTGA GTTTTATTTA GACTTTCTCT TCCTCAACTC
33301 TTCATAAGCT CATGAAAGGT GAAGTAGGGT GCCCTGTGTA TTTATCTTTT
33351 ATATCTGCAG TGCTTAGCAA GTTATAATAA TGCACTTGCC TGGCAAAAGG
33401 CTTTCTCTCA TACATTAGCT TATTTCTCT TCACATTGGC TCTTGTAGT
33451 AATAGGATGC TATTAGTTAT TTTCAATGAG AGAAAGCTAC TAAGAGAAGT
33501 TGTCAGCTA GTGACAGTAA GTGGCTGATA AAGTGAGCTG CCATTACATT
33551 GTCATCATCT TTAATAGAAC TAAACACATA CTGAGTTCT ACTATATTGG
33601 GTCTTTTTTT TTTTTTTTT TTTTTTTTA GAGACGGAAT CTTGCTCTGT
33651 TGTCCAGGCT GGAACGCAGT GGTGCAATT TGGGTCACCA CAACCTCCGC
33701 TTCCCAGGTT CAAGCGATTCA CCTGCCTCA GCCTCCTGAG TAGCTGGAC
33751 TACCAAGTCA CGCCACCACG CCCGGCTAAT TTTGTATTT TTAGTAGAGA
33801 CAGGGTTTCA CCATGTTGGC CAGGCTGGTC TTGAACTCCT GACCTTGTA
33851 TCTGCCGCC TCAGCCTCCC AAAGTGTGG GATTACAGGT GTGAGCCACC
33901 GCGCCCTGCC TATATTAGGA CTTTTATATA AGCTATCTCT AGCTAGCTAG
33951 CTAGCTAGCT ATAATGTTTT TTGAGACAGA GTCTGACTCT GTCAACCCAGG
34001 CTGGAGTGC GTGGCGTGT CTCGACTCAC TGCAACCTCC ACCTCTGGG
34051 TTCCAGTGTAT TCTCCTGCC CAGCCTCCCG AGTAGCTGGG ATTATAGGTG
34101 CATGCCACCA CGCCCAGCTA ATTTTTGTA TTTTTAGTAG ACCAGGTTTC
34151 ACCATGTTGG CCAGGCTGGT CTCGAACCTCC TGACTTCAAG TGATCCACCC
34201 GCCTCGGCCT CCCAAAGTGC TGGGATTATA AGCATAAGCC ACTGTGCCA
34251 GCTGCTCTCT ATATTTTAA TACATATTAT TTCCATTAAT TTTCACAGCA
34301 GTTCATTTTA TAGATGAGGA AACTAGGCCA GAGAAGTAAA ATATCTGCC
34351 CAAGATGATG TAACTAGTAA GTGGCAGGAT CAAGATTCAA ACCAAGCAAT
34401 GTTCAACCT CTGGAAAGCA AGAATGTGGC CACTGTGGAA GGTGCAAGGC
34451 CTTGACAAACA AGAATAGGGA AAAGAAGGAA CTAGAAGGAA AGAGATGGCA
34501 TGGGCTCAGC AGGCCAGGG A GCTCTTAGCT GTGTGTGTG GGAAGCTCAG
34551 AAGGGAGGAA GAGGTTGTCT GTGCAGGTAA GTCTGAGAA CACACCAGAC
34601 TTTTGAGAGG TGGAGCTTCAGGCCAGGTC ATTAGGGGAG AAGGGAGCTA
34651 TAGATTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTAG AGACGGGGTC
34701 TTACTATGTT GCCCAGGCTG GTCTTGAAC CCTGGGCTCA AGTGATCCTC
34751 CCACCTCAGC CTCCCAAAGT GCTGGGATTA GAGGCATCAG CCACCCGCC
34801 CAGCGAGCTA TGATCTAAC ATGTACATCT TACACAGTGC TAATAGAATG
34851 TTGGGTTCT TCCCCAATAT TTTATTTGA AAAAAAATTC AAATATATAG
34901 AAAAGTTGAA AAATGTAGTT CAAAGAACAC CTACATACCT TTCACATAGA
34951 TTCATGATTT GTTAATGTTA TGCCACTTGT TATATATCTC TCTCCCTCCT

FIG.3-14

35001 ATCTGTATACTTTTATTTGC TGAACATTTCAGAGTAAC
35051 TAAAGGCATCTTGATTTAC CCTTGAAACAG TTCAATATGT TTCTGCTAAG
35101 AATTCTCCTA TATAAGTCAG ATATCATTAC ATCTAAGAAA ATTACGGCA
35151 ATTTTACAAT ATAATATTAT AGTCCAAATC CATATTCCT CAGTTGTTCC
35201 AAAAAATGTT CATGGCTGTT TCCTTTTTA ATCTAAATT GAATCCAAGT
35251 TTGAGGCATT GTATTTGGTT GCTGTGTC TAGGGTTTT AAAATCTGTG
35301 CCTTTCTTC TCCCCATGAC TTTTAAAG AGTCAAGACC GGTTATTCTT
35351 ATAGAATAAC CCACATTCTA GATTTGCCTG ATTAGTTTT TTATACTTAA
35401 CGTATTTTG GCAAGAACAT TACATTGGTA ACGCTGTTGG TGATGGGTCA
35451 GTTTGAAGA GTGGAGATGA TTAAACTGCT TTTGTTCAATT GAAGTATCTG
35501 TCAAGACCAAG AGATCCTAA CTGGTGCCT AAATAGGTT CAGAGAACCC
35551 TTTATATATA CACCTGTCC CCCACCTAAA TTATATACAC ATCTTCTTTA
35601 TATATTCAATT TTCTAGGGG AGGCTTCTG GCTTTATCA AATTCTCAGA
35651 GGGCCCCAAG ACCCAAAGAG GTTATGAAAC ACTAGTCTGT CCACTGAGGC
35701 AGGCAACACA GAGCTGGTT CTGGGGCCTT GTTCAGTCTG AACAGCTTC
35751 CCTGGGGAG ATAGCACAAG GCTGTAACCT TGCCCCATCT TGGCTTTGGA
35801 TCAAAGAGGA CTGTCCATT TGTTGTACATA CCTAGGAACC AGGGACAGCT
35851 TATGTGGCCT GGTCCAGGG ATCCAGGAGA ATTCAGTTC TTGTCTTGC
35901 TTTCAGGTGT TCAGAAATGCC AGGATTCCCT CACCAACTGG TACTATGAGA
35951 AGGATGGAA GCTCTACTGC CCCAAGGACT ACTGGGGAA GTTTGGGGAG
36001 TTCTGTATG GGTGCTCCCT GCTGATGACA GGCCCTTTA TGGTGAGTGA
36051 ATCCCTTCAT ATCTGCCCT CTTGGTCTTC AGAGTCCATT GACAGTGCTT
36101 CCAGTTCCCT GTGGCCTGTT AATCTTTAG TCTTCCATC AGCCAGGGCA
36151 TCTCCCTTTA TTTATTCAATT CATTCAACTA GCAGGTATCA ATTGAGCACCC
36201 TACTAAGTGA AAGGTAAGAT CCTTCCCTCA AAGACTTAAT AGTTGAACGT
36251 TGGGAGTGGG AGGAGAGGCA GGCAGAGAGG AGACACAATA TAGTTGGATA
36301 AGGACCTCCA AGGAGAGTGT TACAGGCTGA GAGGAGGATA TACTTAGGTT
36351 GTCTTAGGG AATCAGAAAA GGAGACTCTG GAATAGGCTG GCAGAGAGAG
36401 GGGCTACCTC CTATACCTGC TCTGGACAAA CGACTTTAAG CATACTGACA
36451 GATTTGCCAA CCCTGTATTG GAAGAACTGA TCTTTTTAG TGGGGATGAT
36501 TACTTCTGGG GATTCTTCT CATAACTGAG ACCAAAACAG TTTTGTGCAG
36551 TCTCAGAAAT GACAGGAGGT ACCAATCTGA CACTTCCCTT GGAAGCTCTA
36601 GGGCAGAGAG TGAAAGAGTG GATTTTGACG GGGGCCTTGC TTGGAGGTCA
36651 TTCACCCACC CCTGTCTCA CTCCAGCAAC AGTGATAACT CACTTCTTC
36701 CTCCCTTGT ACACCCCTCT CCCCACCTGC TCACAGGTGG CTGGGGAGTT
36751 CAACTACAC CCAGAGTGT TTGCTGTAT GAGCTGCAAG GTGATCATTG
36801 AGGATGGGA TGCAATATGCA CTGGTGACG ATGCACCCCT CTACTGGTAA
36851 GATAGTGGTC CTTGTCTAT CCTCTCCCAT ATAAGAGTGG CTGGCGGGGA
36901 GGGACAGTGG CAGGGTGAGT TGGGCAGAAG GAGTGTAGG GTAGTCAGAG
36951 CATTGGATTCTTACACACAGC AGTGCTCTTA ACCAGCTCTT TAACTTGAA
37001 GCAGAAATGAT TTACACATGT CTCTACCCCTT TTTCTTACCAACCTTGAA
37051 ATGTCTTCAC TCTGCCCTGC AATCCTCCCA GTGGGAGGCA CTCTTCAGG
37101 ACGATCCCAG AACATTAAG TCAAAGACCC CTTAGAGCTC ACCCTGTCCA
37151 ACCACCTTGG TTGATAAAAG AAGTCAGCCT GGGGCCCATG GAATAGAATA
37201 GTACAAGGGC AAGGTTCTCA TTGTGAGTCA AAGGTAGAGT GAAGAGAACCC
37251 CAGACCATCT CACCCCAACC CAGGCCAGTG TTTTCCAAA TATACCACTT
37301 GCTGCAGATC TAGCTCAGCA CCCCCAGTCC CAGCCCACCC TGAGAACCC
37351 GGCTCCTCAT TCTGAGCAGC CAGCTAGAAT CATGACAAAG AGGGTGGTAG
37401 TGAGACTATG GGTACTGTTG CTTAAAGCCA CATGGTGCAG TGGTTGCTGG
37451 GGGGCTTCTG TGTGGGACTC TAGCATCTTA TTCCCCCTG TGCCCTCTCC

FIG.3-15

37501 CCAGTGGGAA GTGCCACAAT GAGGTGGTGC TGGCACCCAT GTTTGAGAGA
37551 CTCTCACAG AGTCTGTTCA GGAGCAGCTG CCCTACTCTG TCACGCTCAT
37601 CTCCATGCCG GCCACCACTG AAGGCAGGCG GGGCTCTCC GTGTCCGTGG
37651 AGAGTGCCTG CTCCAACATAC GCCACCACTG TGCAAGTGAA AGAGTAAGTA
37701 TTTTGAGAAC CCTTCAGCAG GGGTTCTTGA GCAGAGTCTG TAAATGGGCC
37751 TCAGAGGGCT TAGACCTCCA AAGTCTCATG CAGAACTCCC TTTATTCTCA
37801 TCTCATATCT TTCTCCTGGA CCCCCACTATG CTGTAACCGT ACCTGGGCCT
37851 TGGCACTTAC TGTTCCTCT GCCTCAGGCTA CTTCTACCC GATACTTAAG
37901 GCAAGAATCA CTACACCTTC AGGTGTCAGG TTTCAGGTCA TGTTTGCTCT
37951 TTGAAATCAT CTGGCTTGAT TATGTGTATT AGTTGTTTAT CTTCTATCCC
38001 CTCCACTAGA ATGTAATTG CAGAAGAAC TTGCTGTCTT ATTCACTGCT
38051 GCATGCCAG GGCTTGGAAAG AGTACCTGGC ATATAGTAGG AGTTGATTGA
38101 TTATTATTTT GTCAAGTCAG AGAATGAATG GAGAAAATGT GGTCCATGGC
38151 CCAAAAGAAG TTAAGACCCCT ATCCTAGATT CAGGCCAGAG ACCAGATGGA
38201 GAAAGAGTCT GTGTCTATCT AATACCACTA ATGTCGTACC TCTGGCCGCT
38251 TACCATGTAA ATATTGATTG TGTATCTACC ATGTGTGGAA CACTAGGCTA
38301 GTGCTTGCAC AGCAGGTGAA AGATACTAGA GTTGGGAAG TCAGGAGGAG
38351 CTAAGGTCTG TTCTACAAACC TTATTAGATG AAGAGGAGAG GGAATTGTGT
38401 TCAGGGCAGA GGGAGAAGCA TTCTCCAAA AGTAGGAGTC TTAATCATGT
38451 CTGATGTAGG TTGAGTGTGG CCAGAAAAGG GGCTGTTAAG TATAGAGGGC
38501 CTGGATTATG AAAATCCAGC AGATCCATTG AGAGTTAAG CAGCAAGGTG
38551 TTGTGACCAA GTTAACATT TAGAAGGATC ACTGGTATGG AGGTTGGATT
38601 GGAGAGGGGA AAGCCTAAAG GTATAGAGAC TAGTTAGGAA GCTATTGTAG
38651 GCTGGGCATG GTGGTTCATG CCTGTAATCT CAGCACTTTG GGAGGCTGAG
38701 GTGGGAGGAT TGCTTGAGGC CAGGAGTTGA AGACCAACCT GGCAACATA
38751 GCAAGACCCC GTCTCTGTT TTCTTAATTA AAAGAAAAGT CCAGACGTAG
38801 ACATAGTGGC TCACGCCTGT AATGCCAGCA CTTTGGGAGG CCAAGGTGG
38851 CAGATTGCTT GAGGTCAAGA GTTGGGATT AGGCCAGGCG CAGTGGCTCA
38901 CGCCTGTAAT CCCAGCACTT TGGGAGGCCG AGGTGGCGG ATCACAAGGT
38951 CAGGAGATCA AGACCACCTT GGCTAACACA ATGAAACCCC GTCTCTACTA
39001 AAAGTACAAA AATTAGCCGG GCATGGTGGC GGACGCCTGT AGTCCCAGCT
39051 ACTCGGGAGG CTGAGGCAGG AGAATGGCGT GAACCTAGGA GGCGGAGCTT
39101 GCTGTGAGCA GAGATCACGC CACTGCACTC CAGCCTGAGC GACAGAGCGA
39151 GACTCCATCT CAAAAAAAGAA AAAGAGTTTGG GGATTAGCCT GGCAACATG
39201 GCAAAACCCC ATCTCTACAA AAAGTACAAA AAAATTAGCT GGGTATGGTG
39251 GTGCGGCCT GTAATCCAG TTACTCAGGA GGCTGAGGCA TGAGAATTGC
39301 TTGAGCCTGG GAGGTGGAGG TTGCAGTGTAG CCCAGATCAT GCCACTGCAC
39351 TCCAGCCTGG ATGACAGAGT AAGATGCAT CTCAAATAA AATTAAAAAC
39401 AAAGTTAAA AAAAAAATAG AAGCTATTAC CGTGATCCAG GTAAAGAGATG
39451 TGAATAACTA CAATGATGGA AAGAAGGCAG AGTTCTTAGA GATGGGAGTA
39501 GGAGAGATGA GGGAACTCCA GATTGGGAAG ATGATGTTCA AGTTTCTGGC
39551 TTAGGCCACA GGGTGAGTGG CAATCCCTT CACTGAGATG GGGCATCCTG
39601 GAAAAGGTGT TGCTTTCTG TGTGGGTATC CTGGGCCCT TAGGGGCCAC
39651 TGGTGGCCTG GGACCTGGTA AACCTTCCT GCACAAAGCAG AATTGGTCAA
39701 GCAGGTTTTT AGGACATCTT TACCCCTGCCT CAACTCTTGT CTGGCCCAGG
39751 GTCAACCGGA TGACACATCAG TCCCAACAAAT CGAAACGCCA TCCACCCCTGG
39801 GGACCGCCTC CTGGAGATCA ATGGGACCCC CGTCCGCACA CTTCGAGTGG
39851 AGGAGGTAGA GTGTGTGTCT AATCTGTTTG GTGAGGGTGG GACATGGAAC
39901 AGATCCTCTG GGAAATCAGG CTGTAGCCTT TACCTTTCC TACCCCCAGC
39951 CCATCTCTT GTCTTAGCAT TGAGCCTGTG ACCACTGGTG ACCTATTCA

40001 GCGTAACAGG TTCCCAGGGT AGCAGGGATG GTTGATGGAC GGGAGAGCTG
40051 ACAGGATGCC AGGCAGAGGG CACTGTGAGG CCACTGGCAG CTAAAGGCCA
40101 CCATTAGACA AGTTGAGCAC TG GCCACACT GTGCCCTGAGT CATCTGGTT
40151 GGCCATGGGT GGCCCTGGGAT GGGGCAGCCT GTGGGAGCTT TATACTGCTC
40201 TTGGCCACAG GTGGAGGATG CAATTAGCCA GACGAGCCAG ACACCTCAGC
40251 TGTTGATTGA ACATGACCCC GTCTCCCAAC GCCTGGACCA GCTGC GGCTG
40301 GAGGCCCGGC TCGCTCCTCA CATGCAGAAT GCCGGACACC CCCACGCCCT
40351 CAGCACCCCTG GACACCAAGG AGAATCTGGA GGGGACACTG AGGAGACGTT
40401 CCCTAAGGTG CCACCTCCCA CCCTGGCTCT GTTCTGTCT ATGTCTGTCT
40451 CTCGGATGAA GCTGAGCTGG CTTTCAGAAG CCTGCAGAGT TAGGAAAGGA
40501 ACCAGCTGGC CAGGGACAGA CTATGAGGAT TGTGCTGACC CAGCTGCC
40551 TGTGGGGATC ACAGTTACA GCCAGAGCCT GTGC GGACCC AGCTGTCTGC
40601 CAGGTTTCCT TAGAACCTG AGAGTCAGTC TCTGTCCACT GAACTCCTAA
40651 GCTGGACAGG AGGCAGTGAT GCTAAACCTC GAAGGGCAAC ATGGCCTATG
40701 GAGAAAGCAT GGAGCTCAGA GCCTGGAGTA CGGGCACAGA TAGGATTGAA
40751 TAAATTGTGT AGAAAGACTT TGAAAACAAT AAAGCAAAAG ATGAATGAAC
40801 GTTTTTTTTA GACTTGAGGG ACCAACAAACC CCCAAACCCCC AGATTCTGCC
40851 AGGTCCATGG GGAAGGGAGAA GTTGCTTGA GTGGAGGCC CAAGTAGGGA
40901 GACTTACAGA AAAGAAGTCAGA AGAGCACTGG CTCCCAGGCA GAAATACTGA
40951 TACCCCTACTG GGGCTTCAGG CTGAGCTCT CCCTTCACAA ATCACTTCAT
41001 CTCTCTGAGC CTGTTCTGC ATCTGTGACA TAAGATGGTA AGATAAAGGT
41051 GGCTGTCTCA CCAATTATGT AAGGATTTAA TGTGGAAAAG GACATAAAAGT
41101 TGTATAGTGC TGCCATAGGG ACAGTGTCTA GTAAACGTGA CACATTCTTA
41151 GTATCACTAA GAATCAGGTT CTTGGCCAGG CACC GTGGCT CATGCC
41201 ATCCCAACAC TCTGGGAGGC CTAGGTCGGA GGATGGCTTG AACACAGGAG
41251 TTTGAGACCA GCCTGAGCAA CATAGTGAGA CACTGTCTCT ACAAAAAAAA
41301 AATAATAATA ATAATTGTTT TTAATTAGAT GGGCAGGGCA CTGTGGCTCA
41351 CACCTGTAAT CCCAGCACTT TGGGAGGCCA AGGCGGGAGG ATTGCTTGAG
41401 GCCAGGAGTT CAGGAGCAGC CTGGGCCACA TTCTGTCTC TACAAAGAAT
41451 AAAAAAGTTA ACTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCA
41501 AGAGGCTGAG GAGGAGGATT GCCTGAGGCC AGGAGTTCAA GACTGCAGTG
41551 AGCCTTGATC ACACCACTGT ACTACAGCTT GGGCAACAGA GTGAGACCTT
41601 GTCTCCAAAA AAAAAAGTTT GTTTTTTTT ATCCACTCTC CTCACCAAAAC
41651 AAACTGAGTA AGTTAGAGCC CTCTCAGCTG GCATGTGTTG GAAACAGTGC
41701 CCTCTCATTA AAGTGTGCC CTCACTCCCC TTGCTCTTG GCCTTGGTCA
41751 GTATGATGAA ATTAGTGGGA GGCAGGGCAA CAGAGGGCAG GGAAGAGCTA
41801 GAAATCCATG GCCTGGAAAA GGGAAAGATT TGAGTGGCC AGGTATCTGT
41851 AGAGCCACCA TGCA GAGGAG GGGGGCAGCT AGCCTTGTGT GCTCTGGTGG
41901 GCATGGTCAG CAGGAGGCAG AGCAAAAGGA CAAGGGTAAG TAAACCTGT
41951 GGTGGGACCA AGCCAAGAGC CATCCAGCGT CAGTCTCTC TGGGTAGGCC
42001 AAGTAAAGCA GGAGCATACC CCAGAGAGAA AGTTCGCAGG GCTGTTCA
42051 TGCA GTGCTGCTG TGGA CTTCAAC CCTTCTTGTT CCTTCTTCAG TAAGTGA
42101 TAACAGTCAT TGACCATGAC TATTATCGAC CGCTTTGAA AATGTAAACA
42151 TAGTGA CTTT ATTGCTGTAA AAATCATACG TGTTTATCAT CTTAAATTC
42201 AGGAAACATG GACAGGTACA AAGATGTGCA AAATATCATC CAAAATCCC
42251 TTTGCTGGCC AGGCACGGTG GTCACGCCT GTAATCCCAG CACATTGGGA
42301 GGCGGAGGCG GGCAAATCAC TTGAGGTCAG GAGTTTGAGA CCAGCCTGGC
42351 CAACATGGTG AAACCCCTATC TCTACTAAAA ATACAATAAT TAGGCTGGGC
42401 GCAGTGGCTC ACGCCTATAA TCCCAGCACT TTGGGAGGCC GAGGTGGGCG
42451 AATCACAAGG TCAGGAGTTT GAGACTAGCC TGGCCAATAT GGTGAAACCC

42501 CATCTCTACT AAAAATACAA AAATTAGGGC CGGGTGTGGT GGCTCACGCC
42551 TGTAAATCCA GCACTTAGGG AGGCCGAGAC AGATGGATCG CGAGATCAGG
42601 AGTTGGAGAC CAACCTAGCC AACATGGTGA AACCCCATCT CTACTAAAAA
42651 AATACAAAAA TTATTCGGTT GTGGTGGCAC ACGCCGTAA TCCCAGCTAC
42701 TTGGGAGGCT GAGGCAGGAG AATCTTGTGA ACCTGGGAGG CAGAGGTTGC
42751 AGTGAGTGGA GATCCCAGCG TTGCACTCCA GCCTGGGCGA CAGAGTGAGA
42801 CTCCATCAA AAAAAGAAAAA AAAAAAAA AAATTAGCCG GGCCTGGTGG
42851 CGTGCACCTA TACTCCCAGC TACTTGGGAG GCTGAGGCAG GAGAATCGCT
42901 TGAACCTGGA AGGCGGAGGT CGCAGTGAGC CGAGATCGTG CCATTGACT
42951 TCAGCCTGGG CGACAGAGCG AGACTCTGTC TCAAAAATAA TAATAATAAC
43001 AATAACTAGC CGGGCCTGGT GGCACATGCC TGTAGTCCC GTTACTCAGG
43051 AGGCAGGAGC ATGAGACTCA GGTGAACTAG GGAGACAGAG GTTGCAGTGA
43101 GCCAAGATCA CACCACTGCA CTCCAGCCTG GTTGACAGAG CGAGACTCTG
43151 TCTCAAAAAA AAAAATCC CATTGCTCA TTTTTGGAT ACTAGTATAA
43201 CTATCACT AAACCAAGTTA GTACTAAAT CAAGCAGATA TGGGAGATGG
43251 TGAATTACCA TCTACAGTGT TGTCATATAT GTCACATACT GAGCATTATC
43301 AGCTAGTAGA ATCTAGTTAA TTGTTCTATG TGTGATGTAT GCAGAGTCC
43351 CATTGGAT GTGTTTAC TATGCTTAA TAAATGACTG ATGTCAGCAA
43401 CCCCCAAATG ATACATCTGA TGTAAGAGCC CCTGTTCCCC AATAATAACA
43451 TCTAAACTAT AGACATTGGA ATGAACAGGT GCCCCTAAGT TTCCCTCCCTC
43501 CAGGGTTCT TGCCGGTCT CTGAGGACTA CACATCCCTA CTCCCGTCTT
43551 TCCTCATCTT CAGGCAGT AACAGTATCT CCAAGTCCCC TGGCCCGAGC
43601 TCCCCAAAGG AGCCCTGCT GTTCAGCCGT GACATCAGCC GCTCAGAAC
43651 CTTCTGTTGT TCCAGCAGCT ATTACAGCA GATCTCCGG CCCTGTGACC
43701 TAATCCATGG GGAGGTCTG GGGAAAGGGCT TCTTTGGCA GGCTATCAAG
43751 GTGAGCCAG GCAACAATTG CTTTGCTTT CTGCCCCAG TCCCTCTGTC
43801 ACTGTCTTTC GGGGATTCT CATCACTTGG CCCCACCCCA CACCATGAG
43851 GATGCCAGGC CTCTTCCCTG GCTTTGGTG TTGGTGTGAG AGGTATCCTT
43901 CACCCCCACC CAGGCCACCT AAGGTCAATG TTGCTGTTAC AGTGAGCTTG
43951 TGGACCTGGA GATCCAGGTT GGGTTGAGCT GTGCTGTGG CCCTCCTGCC
44001 TCCAGTCAGT GGGTGTGGT TAGGTGCCCTG CAGACCTCAG TACCGGGCAT
44051 GCTACAAGGA GCACACAGGG GAATGGCTCC TGCCTCCCTG GTGAACAGTC
44101 TCAGGGACTA ACCTCTCTCT TTCTCTCCCTC CTCCCTCCCTCT TCTGCTGAGA
44151 ACTGGGAGGG GGGGTCAAGT AAGACGTGTG TCTCAGCTTG GGGGCAGCAG
44201 GGCTGGAGAG CTCAACCCCG ATCCACCCAG CTCCCTGGTG CATGTCTTTC
44251 GCACTGACCT TCCCTGCCCTC AGACTTCTGT TCACTCAGGA GACTCACTTC
44301 TATGCCAAAT GACCAGAGCC CCTGCTTGGC TTGGCAGCAT CCCCTCCTGCC
44351 CTTCTCCCTC ACTTCCCTT TCTGGGTTCT TGCCTGTCCCT CTGTGCATGC
44401 CCAGCTCTCC AGGAAAGAGG GTTGTCTTCC GTGTGAGTCC CATGTTGCTC
44451 CACGCTGCAT CTTCCACACA TGAACCTGT CATTCTGACC CGGCTCAGTG
44501 TGCCCTCCAA GGGATGGGAT GGCCAGCTGC ATAGATTTTC TCAAACAGTT
44551 CTCCAGAACT TCCCTCTGGTC TCAGCACCAT TAACAGTCAC CCTCCCTGTA
44601 GGTGACACAC AAAGCCACGG GCAAAGTGT GGTGATGAAA GAGTTAATT
44651 GATGTGATGA GGAGACCCAG AAAACTTTTC TGACTGAGGT AAGAAGATGG
44701 AGGGGGCCCG GGAGGTGGT GTCACCATGG GAAGAGAGAA GACCTTACAA
44751 ATAATGGCTT CAAGAGAAAA TACAGTTTG AATTACTGTC TTAAAGACTA
44801 AGCAGAAAAG AGCCCTAGAG GAATATCCA CTCCCTCTAA ATTACAGCGT
44851 AATTATTTGT TCAATGAACA CTTACTAAA GCAACACAAA CAGGGTACAA
44901 GGGATGCAGT AACAAAAAGAT ACAGGGTCA GAAGAGCTCT CAGGTTATGA
44951 GGATGATGGA CATGAAAACA CTCCAATTAA GTACAACCTCA ATGTTATAAT

45001 CCTCACCTGA ACGCCCTGCT AAGGGAGCCT GGAGGGGAGC TCCCTGAGCA
45051 CTCACACTCC TTGGGCATTT ACAGTTTCA CTACCCCTCC CAAGTTACTT
45101 CATGGAGTAA CTTAAGTTGG GGACACCTGT GGTCTGGGTA TTGCCCTCCA
45151 AGCCACTTGG CCACATCCCAC CCCAGTTCTC CCAATGCAGT TCCAAGGGTA
45201 AGGCCTATGA AGCCATCTCC ATCTATATGG TGGTGGTCTT CCCTCATCCT
45251 GATCTTAGTG CCCTGTCTA TCACAAGATA GGAGGTAGGA GATACAGGTG
45301 GTAACACTTG TCAAGCTGAT TCCTTGGAGG GAAGAGGTAA GGAAGACAGT
45351 GAGAAGTTAA CCACCAAGCTT TCCTTGGCTT CCCCCACCCC CAGGTGAAAG
45401 TGATGCGCAG CCTGGACCCAC CCCAATGTGC TCAAGTTCAT TGTTGTGCTG
45451 TACAAGGATA AGAAGCTGAA CCTGCTGACA GAGTACATTG AGGGGGGCAC
45501 ACTGAAGGAC TTTCTGCGCA GTATGGTGAG CACACCCACCC CATAGTCTCC
45551 AGGAGCCTTG GTGGGTTGTC AGACACCTAT GCTATCACTA CCCTAGGAGC
45601 TTAAAGGGCA GAGGGGGCCCT GCTTGCCTC CAAAGGACCA TGCTGGGTGG
45651 GACTGAGCAT ACATAGGGAG GCTTCACTGG GAGACCACAT TGACCCATGG
45701 GGCCTGGACC ACCAGGTGGGA CAGGGCTCAA CAGCCTCTGA AAATCATTCC
45751 CCATTCTGCA GGATCCGTTCC CECTGGCAGC AGAAGGTCAG GTTTGCCAAA
45801 GGAATCGCCT CCGGAATGGT GAGTCCCACC AACAAACCTG CCAGCAGGGC
45851 GAGAGTAGGG AGAGGTGTGA GAATTGTTGGG CTTCACTGGAA AGGTAGAGAC
45901 CCCTTCCTAT GCAACTTGTG TGGGCTGGGT CAGCAGCTAT TCATTGAGTT
45951 TGTCCTGTC ACTGAAACTG ACCCCAGCCA ACTGTTCTCA GTTCACAGCC
46001 CTGTTTCAA AGAATTACAC ATCTCTAAAG GCAAACAGGG CACGGACAAG
46051 GCAAACCTGGA GAGGCAAACG GTAGCCTGAG ATGGCCTGGG CTTGCCATCA
46101 CAGGTATTCA GGTGCTGAGG GCCCTTAGAC CAACTAGAGC ACCTCACTGC
46151 CTAGGAATC AATGAAGGGG AAATGAGTTC TAGGGAGCC CTGAAGGATC
46201 AGAATTGGAT AAAGTTCTTA TTGGCAGAGA GGCACCAAGGA TTGAAGTGAC
46251 AGGAGCAAAG ACCTGGGAGG AAAGAGGAGA AAATCATCTA TTTCACCTGG
46301 AAACAAATGA TTCCAAGCAT AGAAAATAATA ACAGCTGACA AGTACTGAGT
46351 GCCCTCTATA TGCTAGGCAC TGGGCTGAGG GATTAACATG CATGTGCATG
46401 TTATTCCTC ATGACAACCT TGGTTCCAG ATAAGCTGGA CTGGAAAGGG
46451 ACAGAGCTGG GATCCTGGGC TAATCAGTCT GGTGCCAAG CCTGAGACTT
46501 TAGCCACTGC CCTTCACATG GGGGTCCATG AAAATAGTAG TAGTCTGGAA
46551 CAGTTGGGG GTACATCAAG GTCGCTGTGT TTAAGCTAT GGAGTCTGG
46601 CTATAGGAGA CAAATGTAAA AGAGTTTTT GGTGACTGG CTTTTTGGTT
46651 TTTTTGTTTG TTTGTTGTT TGTGTTGTT TTTGTTGTT TTTCTGTT
46701 TCTGGGGCTT GAATCAGGAA GGAGGTTTTT TTGTTGTTGT TGTTTGAGA
46751 AAGGATATTG CTCTGTTGCC CAGACTGGAG TGCACTGGCA CGATCATGGC
46801 TCACTACAGC TTGACCTCC TGGGCTCAAG CAATCCTCCT GCCTTAGCCT
46851 CCCAAGTAGC TGGACTACAG GTGTGTACCA CCACACCTAA TTTTTGAAT
46901 TTTTTTTCT TTTTTTTTT TTTTTTTTTT GGTAGAGACA GGTTCTCACT
46951 TTGTTGCCCA GGCTGAATC TCAAACCTCT GGGCTCAAGC ATTCCCTCTG
47001 CCTCGCCCTC CCAAAGTGTGTT GGGATTACAG TTGTGAGCCA CCATGCCCG
47051 CAGGAAAAGA TTTTAAGCA AGAAAGCTTA AGAGCTGTGG TTTTTCAAA
47101 ATGAGTCTGG GCTGGCACAG TGGCTCATGC CTGTAATCCC AGCACTTTT
47151 TGGGAGGCCG AGGTGAGTGG ATCACTTGAG GTCAGGAGTT TGAGACCAGC
47201 CTGGCCTACT GGTGAAACCC CTGTTCTAC TAAAGAAAAA AATGCAAAA
47251 TTAGCTGGC GTGGTGGTGC ACGCCTGTAG TCCCAGCTAC TCAGGAGGCC
47301 GAGGCAGGAG AATAGCTTGA ACCTGGGAGG CAGAAGTTGC AGTGAGCCAA
47351 GATCACACCA CTGCATTCCA GCCTGGGTGA CAGAGTGAGA CTTCATCTCA
47401 AAAAAAAA AAAAGAGAGA CTGATATGGT TAGTACATTG GGGTGGAAATG
47451 CGGAGGGTCC AGGGAATGGA GCCCTGCATA GGGGCTAAT GAAACATTTC

47501 AGATTTCTGA ATTAAGGTAG TGGCTGTGGG GACAGGAGCC TGGGAGGCAG
47551 GGTGGAGTCA GAATGGAGAG ACTGGTGGC AATGAGGGAA CAGGAGGAGG
47601 AGGAGGAGGA GTTACGAGTG GCTTGAGGTG TCACCTTACCA GACATTTGGG
47651 GGATGGGGGA TAGCCGTGAT TGTTGAGCAA CTGGTTGGG AAGAGCTAGC
47701 ATTGATCCCT GCTGTTCTG GCTAGCAGAA CCTATCAGCA TCTTCTGGC
47751 AGGAAACTGG CTCCATGAGA CTGGCTTAGG GAGAGGCTGC TAGTCACCTA
47801 ATCTGCAGAG AAGGGGCAGC TGGAGCTGTG GGACAGAAGA GGCATCCATG
47851 TAGCTGGTGG GGGGTGCTCA GCTTGTGAAG AGGAGATGGC TTTGAGCAGG
47901 GCTGACACTG AAAAGGCTGG AAGAAAAAAA CAGACACACA AGAGTCTCAG
47951 GATCAGGTAG CATAGGAAAG TTGTGGACAG TCTTTGAGGA GCACTCCCTC
48001 AGGCAGGCAG GCAGGCAGGT CATGAGCTAT AGCGATTCAAG GAAGAGCTCC
48051 CTGGGTGTGT GAGCAGCTCC AGGAGCCTAA GGGATGAAAG TAGTATTGCA
48101 GGGGGCTGGA GAGCAAGGAG TGGCTCCTTC TACATTGCA AGGGAAAGGAG
48151 AAAGGAAGTT GCTCTGAGA GTGGTAAGAG TCAGTGGTGG AGGCCTGGAG
48201 AGGAGACATA ACAAAACAAAT TTGTTGACAA ACATTTGGT AGGAAGGGGG
48251 AGAGCTTAA GTTTAGACAG TGGGGAAAGT GGAGTCTTAG AGGAGGTGAA
48301 TGTCTGAAAG ACAGAGCTAG CTGGAGCAAG AAGTCACCTC TCTGTTGCAG
48351 GCAGGAAGGA TCCAAAGTGG CTCAAGCCAG AGATTGGGAG AGTGGGGAGG
48401 AGGGAGCAGC CTGGATCTAA GTAAAATGGG TAGAGGTGGA GGGGGTGCTG
48451 CAACGGCCAG GGTTCCTGTA AGTTGGGAC ATTAGGAGAG AGCTGTGAGG
48501 GCTTTGCCA GCCACTGTGC TAGTGAATTGG TGAACCAAAG GATGGGCAGG
48551 AGATGGCAGC AGGGAAGCAG AGGAAGTCCA GGCTTCCTGT TGGTATTGGG
48601 ACAAGGGAGA GGCCATAGGA GGCCCTGGCC CTGTTGTCCA GTTTGGGTTC
48651 TGAAGCTGGG TGGCATGGC CTGGTAGGAG AGCATCTATG GCGCCCAATT
48701 CCAGATTCAAG GGTCTAGTTG ATTTGCTGGC CCTGTAGCCT CAGCTCATGC
48751 TTCTGTTCCA GGCTTATTTG CACTCTATGT GCATCATCCA CCGGGATCTG
48801 AACTCGACA ACTGCCTCAT CAAGTTGGTA TGTCCCCTG CTCTGGGCCT
48851 GGCCTCCAGG GTCCATCCT TCCTGGCTTC CTTGTACAA AGGAGGCTGA
48901 CTTGTCCCCCT CTGGCTAGAG GGCAGAGGTG TTGCCTAGGA GCTCCTATCT
48951 TTCCCTTCCCT GCTTCTTCCA ATGCCCTTCT CTGTCCTCTG GGAGCTCCGA
49001 GACACACACA GACATAATTTC CACCTTCTCT CATTAGCAAC CTTTGAAATA
49051 ATTTGATTAG AAGGGACTTC AGAAGTTTG TGACTATATG TAGAAAACCC
49101 TGTCAATTAA CCTGCTTTTG CCCCCATAGTA GTCTTGAAA ACAGTTCAATT
49151 GCTGACCCCA TTTTACAGTG GTGGCACCTG AAGCCTCAGC CTGAGGCCAC
49201 CGAGCTAGTA AATTACAGG GACCAGTTTG AGACCAGCAT TCCTCCACT
49251 GCCCCTCAGC TGTGGTGGTT ACAATGTTGT TTGCTTACT GACTTGCTAT
49301 CTGGCTTCCCT GGGGTGTCTAC CGGCTGGCCC TGGCTCTGCC CTCTAGACCC
49351 ACACCACGCA ATCTTCATTC CTTTCCCACA TGACTGCCCT GTAGCTATTTC
49401 AAAGAGCTTG TCTCCCCAA GTCTCCCAT CTACTGCCTC CACCTTGCCT
49451 TTTTCTGTCT TATCCTGGTT CTAGCCACTG CCTGAAATCA TTTTAGGAAT
49501 AAGACAGGAC AGGGAAAAAC AAAAGCAACC CCCTGTCCCCA CCTCTGAGTT
49551 CCACTCTCCA AGTCCCTGAG CCTCACCTCC AGGGCTCCAG TGGCTCTGCC
49601 ATGAACCCAC TGTGGGCTGG GAGTCTGCTG TGCACAGATA CCAGACCCCTC
49651 AGAAACACAA ATGCCAAGTG TGTCTGTTTT TTTGTTTGT TTTGTTTGT
49701 TTTTCTGTCT GAGTCTCATT CTGTTTCCCA GGCTGGAGTG CAGTGGTGCA
49751 ATCTTGGCTT ACTGCAGCCT CTACCTCCCG GGTTCTAGTG ATTGTTCTGC
49801 TTCAGCCTCC CAGTAGCTAG GACTACAGGC GTGTGCCACC ACGCCCAGCT
49851 AATTTTTTTT TTTTTTTT TGTATTTTA GTAGAGACAG GGTTTGGCCA
49901 TGTTGGCCAG GCTGGCTTG AACTCCTGAC CTCAGGTGAT TCACCCGCC
49951 TGGCCTCCCA AAGTTCTGGG ATTACAGGTG GAAGCCACCG TGCCCTGGCCT

FIG.3-20

50001 GAGTGTGTCT ATTTGATAGA GCTTTCTGCT CTGATTCTCC CTTGCTATAC
50051 ACCTTTCTC CCCTTCTAG TGGCTTCTCT TGCCATGCT TCCTCCCCAG
50101 GGCCAGGTTT GAGAACATCC CCATGAAGTC CTGACCTGTC TTTTATCCTA
50151 CCAGGACAAG ACTGTGGTGG TGGCAGACTT TGGGCTGTCA CGGCTCATAG
50201 TGGAAGAGAG GAAAAGGGCC CCCATGGAGA AGGCCACAC CAAGAAACGC
50251 ACCTTGCAGA AGAACGACCG CAAGAAGCGC TACACGGTGG TGGGAAACCC
50301 CTACTGGATG GCCCCTGAGA TGCTGAACGG TGAGTCCTGA AGCCCTGGAG
50351 GGGACACCCG CAGAGGGAGG ACAGATGCTG CCCTTGCATC AGAGCCCTGG
50401 GAATTCCAGG GGAGGCCCTGT GAAGCGTAGG ACCGGATACC CAGAGCTGAG
50451 GATATTTTC CCTTGCAGG TGGGGCCTCA CGATTAGCT CCTGAGCTCA
50501 GGGGGCTGGG AACTGATCG TGTCCTCATCA TGGGGATAA GGTGAGTTCT
50551 GACTGTGGCA TTTGTGCCTC AGGGATCGCT AAGAGCTCAG GCTATTGTCC
50601 CAGCTTTAGC CTTCTCTCTC CATGGTGAGA ACTGAAGTGT GGTGCCCTCT
50651 GGTGGATAAT GCTCAAACCA ACCAGAGATG CTGGTTGGGA TTCTTGAAT
50701 CAGGGTTGT AGGCCTCAGA AATGGTCTGA ATACAATCCA TTTTGGAGTC
50751 TGAGGCCAG AGAAGTTCTAG TGAATTGCCT AGGAGCATAAC AGCTGCCTAA
50801 TGGCAGAGGC TAGATGAACC CTAGTCTGGT TCTTTCCAC TTTAACGTGC
50851 AGTTTCATCC TAGGCAGTGT TATGTTATAA GGGCTCTCCA AGGCAGTTCA
50901 CCTACGGCTG AGGAAGGACT ATTTTCAGGT GGTGTCCTGCG CAGGACAGCC
50951 TGTGGGGTGT CCCTACAGAA CCTGTTCTAG CCCTAGTTCT TAGCTGTGGC
51001 TTAGATTGAC CCTAGACCCA GTGCAGAGCA GTAAAGGGAT GTAAACTTAA
51051 CAGTGTGCTC TCCCTGTGTT CCCAAGGAAA GAGCTATGAT GAGACGGTGG
51101 ATATCTTCTC CTTGGGATC GTTCTCTGT AGGTGAGCTC TGGCACCAAG
51151 GCCATGCCCG AGGCAGCAGG CCTAGCAGCT CTGCCCTCCC TCGGAACCTGG
51201 GGCATCTCCT CCTAGGGATG ACTAGCTTGA CTAAAATCAA CATGGGTGTA
51251 GGGTTTTATG GTTTATAACG CATCTGCACA TCTTGCACAC GTTCGTGTTT
51301 CATTGGTCTT AAGAGAAGGA CTGGCAGGGT TTTTTGGTT TAGATGGAGC
51351 CTCACCTCGT TGCCCGAGGCT GGAGTGCAGT GGCACAATCT GGGCTCACTG
51401 CAACCTCTGC CTTCTGGGTT CAAGTGAATC TCCTGCCTCA GCCTCCCAAG
51451 TAGCTGGGAC TACCGGCACA CACCACCATG CCCGGCTAAT TTTTGTATTT
51501 TTAGTAGAGA CAGGGTTTCA CCATGTTGGC CAGGCTGGTC TTGAACCTCG
51551 GACCTCAGGT GATCCGCCTG CCTCAGCCTC TAAAAGTGT GGAATTAAATA
51601 GGCCTGAGCT ACCTCGCCCG GCCAGGTTTT TTTTTTTTT TTTTTAGTTG
51651 AGGAAACTGA GGCTTGAAG AGGGCAGTGG CTTGCACATG GTCGATAAGG
51701 GGCAGATGAG ACTCAGAATT CCAGAAGGAA GGGCAAGAGA CTGTTCATGT
51751 GGCTGTCTAG CTAGCTCTG GGCCAAATGT AGCCCTTCTC AGTTCCCTTC
51801 AAGTAGAAAGT AGCCACTCTA GGAAGTGTCA GCCCTGTGCC AGGTACCAACG
51851 TGGACAGAGT GAGGAATCTT GGAAAGATTG CTACCTTCTAG GAGTTTAGTC
51901 AGGTGACAGC ATATCTCAGC GACTCAAACA CACACACATT CAAAGCCTTC
51951 TGTAATTCTC ACAAAAGTTGT GAGGGGTAGA GGAGAGGAGA GACAAGGGAT
52001 GGTTAGGATA ATGAAGGAAT TTTTTGTTT TGTTTTGTT TTTGAGATGG
52051 AGTTTCACTC TGTCACCCAG GCTGGAGTGC AGAGGTGCAA TCTTGGCTCA
52101 CTGCAGCCTC CGCCTCCCGAG GTTCAAGCAA TCCTCCTGCC TCAGCCTCCC
52151 AAGTAGCTGG GACTACAGGT GTGCGCCACC ACGCCCTGGCT AATTTTGTA
52201 TTTTCAGTAG AGACAGGGTT TCGCCATATT GGCCAGGCTG GTCTCAAATG
52251 CCTGACCTCA GGTGATAACAC CCGCTTCAGC CTCCCAAAGT GCTGAGATTA
52301 CAGGCATGAG CTACCGTGCC TGGCCATGAA GGAAGATTTG TTTTAAAAAA
52351 TTGTTTTCTT TAATATTAAT TGAACACCTC TGTCAGAGC ACTGGGCTGG
52401 TGCCAGAGGG TTTCAGACAT GAATCAGATC CAGCACCTCA TAGAGCCTTA
52451 ATCTGGCACA CACACACAGC CACAAGGAGA CACAGACAAG GCAGGGTAGG

52501 ATGAGTGGAA GCTAGGAGCA GATGCTGATT TGGAACACTT GGCTTCTGCA
52551 GTGAAGCCCC TTCTTAGTCC TCTTCAGTAA CCCAGCTCTC AGTGGATACA
52601 GGTCTGGATT AGTAAGATTG GGAGAGATGA TTGGGGATTG GGGAGAGCTC
52651 TCTAACCTAT TTTACCACCT CCTCTTCTGC CATTCTTCCT GTCCACATCC
52701 CCAGCATCCC TTTCCCTTGC CAAGTATCTG TGGCCTCTGT AGTCCTTGT
52751 AAACAGCTGT CTTCTTACCC TACAGATCAT TGGGCAGGTG TATGCAGATC
52801 CTGACTGCCT TCCCCGAACA CTGGACTTTG GCCTCAACGT GAAGCTTTTC
52851 TGGGAGAAAGT TTGTTCCCAC AGATTGTCCC CGGGCCTTCT TCCCGCTGGC
52901 CGCCATCTGC TGCAAGACTGG AGCCTGAGAG CAGGTTGGTA TCCTGCCTT
52951 TTCTCCCAGC TCACAGGGTC CTGGGACGTT TGCCCTGTGTC TAAGGCCACC
53001 CCTGAGCCCT CTGCAAGCAC AGGGGTGAGA GAAGCCTTGA GGTCAAGAAT
53051 GTGGCTGTCA ACCCTGAGC CATCTGACAA CACATATGTA CAGGTTGGAG
53101 AAGAGAGGAG TAAAGACATA GCAGCAAGTA ATCTGGATAG GACACAGAAA
53151 CACAGCCATT AAAAGAAAAGT TTAAAAGAAG GAAATTCAACC CAAACCATT
53201 GAATACAGTA AGTGTATTCA TCTTTCGATA TTCCCTGTGTC CATATCTACA
53251 CATATACTTT TTTTTATAGT AAATAGTTCT GTATTTGCC CTGCATTCC
53301 CTTGTGTTTA CTATCCAGTC TTCTGTGTTA TCATTTTGT CGACAACATG
53351 AAATTCTATT GAGAGACTGT CTGAACATAT TGTATGTAG ATGTTCAGGT
53401 TTTTCCAGTT TCTCTTACA ATAGGTATTT AACTACAGTG AGCAGTTTA
53451 TGCAATTAGC TAATTTCTCC TTTGAGGAAG TATTTCAAA ATTACCTTTA
53501 TTCTTCTCAG GTATAATTT CATTATTACG AAAGTTACCC TAGGTCTTTT
53551 CAAGTGTGTG GTAAAAAAAC GAGAATCTGG CTGGCGCGA TGGCTCACAC
53601 CTGTAATCCC AGCACTTTGG GAGGCTGAGG CTGGTGGATC ACCTGAGGTC
53651 TGGAGTTCGA GACCAGCCTG GCCAACATGG TGAAACCCCCA TCTCTACTAA
53701 AAATACAAAA CTTAGCCAGG CATGGTGGCA GGTGCGTGTAA ACCCCAGCTA
53751 CTTGGGAGGC TGAGGGCAGGA GAATTGCTTG AACCCAGGGG CGGAGGTTGC
53801 AGTGAGCCGA TATCACGCCA TTGCACTCCA GCCTCGGCAA CAAGAGTGA
53851 ACTCTGTCTC AAAATGGGG TTCTTTCTC GCCATCAAAA ATCATGTTTC
53901 TTTTAAAAAC AAGTTCAAAAC ATTACCAAAG TTTATAGCAC AGGAAATACG
53951 TCTTCTGTAA TCTCCCTTAA CCAATATATC CCTCAACATT CTCTCACCC
54001 CCAACTCCAC CCTCCCAAGGA TAACCAAGTT GGACATAATC TTTATTTAAA
54051 AATGGTTTCC GGATAGAGAA AGCGCTTCGG CGGCGGCAGC CCCGGCGGCG
54101 GCCGCAGGGGG ACAAAAGGGCG GGCAGGATCGG CGGGGAGGGG CGGGGGCGCG
54151 ACCAGGCCAG GCCCCGGGGC TCCGCATGCT GCAGCTGCCT CTGGGGCGCC
54201 CCCGCCGCCG CCCTCGCCGC GGAGCCGGCG AGCTAACCTG AGCCAGCCGG
54251 CGGGCGTCAC GGAGGCGGGCG GCACAAGGGAG GGGCCCCACG CGCGCACGTG
54301 GCCCCGGAGG CGGCCGTGGC GGACAGCGGC ACCGCGGGGG CGCGGGCGTT
54351 GGCGGCCCG GCCCCGGGCC CCAGGCCAGG CAGTGGCGGC CAAGGACCAC
54401 GCATCTACTT TCAGAGCCCC CCCCAGGGGC GCAGGAGAGG GCCCGGGCTG
54451 GGCGGATGAT GAGGGCCCAG TGAGGCAGCA AGGGAAGGTC ACCATCAAGT
54501 ATGACCCAA GGAGCTACGG AAGCACCTCA ACCTAGAGGA GTGGATCCTG
54551 GAGCAGCTCA CGCGCCTCTA CGACTGCCAG GAAGAGGAGA TCTCAGAACT
54601 AGAGATTGAC GTGGATGAGC TCCTGGACAT GGAGAGTGAC GATGCCTGGG
54651 CTTCCAGGGT CAAGGAGCTG CTGGTTGACT GTTACAAACC CACAGAGGCC
54701 TTCACTCTG GCCTGCTGGA CAAGATCCGG GCCATGCAGA AGCTGAGCAC
54751 ACCCCAGAAG AAGTGAGGGT CCCCCGACCCA GGCAGAACGGT GGCTCCCATA
54801 GGACAATCGC TACCCCCCGA CCTCGTAGCA ACAGCAATAC CGGGGGACCC
54851 TGCAGGCCAGG CCTGGTTCCA TGAGCAGGGC TCCTCGTGGC CCTGGCCAG
54901 GGGTCTCTTC CCCTGCCCCC TCAGTTTCC ACTTTGGAT TTTTTTATTG
54951 TTATTAAACT GATGGGACTT TGTGTTTTA TATTGACTCT GCGGCACGGG

FIG. 3-22

55001 CCCTTTAATA AAGCGAGGTA GGGTACGCCCT TTGGTGCAGC TCAAAAAAAA
55051 AAAAAAAAAT GATTCCAGC GGTCCACATT AGAGTTGAAA TTTTCTGGTG
55101 GGAGAACCTA TACCTTGTC CTTTATAGGC CAAGGACCGC AGTCCTTCAG
55151 TAACACCACT GTAAAAGCTT GAGGAGAAAT TGTGAAGCTA CACAGTATT
55201 GTTTCTAAT ACCTCTTGTG ATTCTAAATA TCTTTAATT ATTAAAAAAAT
55251 ATATATATAC AGTATTGAAT GCCTACTGTG TGCTAGGTAC AGTTCTAAC
55301 ACTTGGGTTA CAGCAGCGAA CAAAATAAG GTGCTTACCC TCATAGAAC
55351 TAGATTCTAG CATGGTATCT ACTGTATCAT ACAGTAGATA CAATAAGTAA
55401 ACTATATTGA ATATTAGAAT GTGGCAGATG CTATGGAAAA AGAGTCAAGA
55451 CAAGTAAAGA CGATTGTTCA GGGTACCAAGT TGCAATTAA AATATGGTCG
55501 TCAGAGCAGG CCTCACTGAG GTGACATGAC ATTAAAGCAT AAACATGGAG
55551 GAGGAGGAGT AAGCCTGAGC TGTCTTAGGC TTCCGGGCA GCCAAGCCAT
55601 TTCCGTGGCA CTAGGAGCCT GGTGTTCCG ATTCCACCTT TGATAACTGC
55651 ATTTTCTCTA AGATATGGGA GGGAAAGTTT TCTCCTATTG TTTTTAAGTA
55701 TTAACCTCCAG CTAGTCCAGC CTTGTTATAG TGTTACCTAA TCTTTATAGC
55751 AAATATATGA GGTACCGGTA ACATTATGCC CATTCTCAC AGAGGCAC
55801 CTAGGTGAAG GAGTTGCTT GACGTTATAC AACCAAGGAAG TAGCTGAGCC
55851 TAGATCCCTT CCACCCACCC CATGGCCCTG CTCATGTTCC ACCTGCCTCT
55901 AATTACCTC TTTCCCTTCT AGACCAGCAT TCTCGAAATT GGAGGACTCC
55951 TTTGAGGCC C TCCCTGTGA CCTGGGGGAG CTGGCATCC CGCTGCCTGC
56001 AGAGCTGGAG GAGTTGGACC ACACGTGAG CATGAGTAC GGCCTGACCC
56051 GGGACTCACC TCCCTAGCCC TGGCCCAGCC CCCTGCAGGG GGGTGTCTA
56101 CAGCCAGCAT TGCCCCCTTG TGCCCCATTC CTGCTGTGAG CAGGGCCGTC
56151 CGGGCTTCCT GTGGATTGGC GGAATGTTA GAAGCAGAAC AAGCCATTCC
56201 TATTACCTCC CCAGGAGGCA AGTGGCGCA GCACCAAGGA AATGTATCTC
56251 CACAGTTCT GGGCCTAGT TACTGTCTGT AAATCCAATA CTTGCCTGAA
56301 AGCTGTGAAG AAGAAAAAAA CCCCCTGGCT TTGGGCCAGG AGGAATCTGT
56351 TACTCGAAC CACCCAGGA CTCCTGGCA GTGGATTGTG GGAGGCTCTT
56401 GCTTACACTA ATCAGCGTGA CCTGGACCTG CTGGGCAGGA TCCCAGGGTG
56451 AACCTGCCTG TGAACTCTGA AGTCACTAGT CCAGCTGGGT GCAGGAGGAC
56501 TTCAAGTGTG TGGACGAAAG AAAGACTGAT GGCTCAAAGG GTGTGAAAAA
56551 GTCAGTGATG CTCCCCCTT CTACTCCAGA TCCTGTCTT CCTGGAGCAA
56601 GGTTGAGGGA GTAGGTTTG AAGAGTCCCT TAATATGTGG TGGAACAGGC
56651 CAGGAGTTAG AGAAAGGGCT GGCTTCTGTT TACCTGCTCA CTGGCTCTAG
56701 CCAGCCCCAGG GACCACATCA ATGTGAGAGG AAGCCTCCAC CTCATGTTT
56751 CAAACTTAAT ACTGGAGACT GGCTGAGAAC TTACGGACAA CATCCTTTCT
56801 GTCTGAAACA AACAGTCACA AGCACAGGAA GAGGCTGGGG GACTAGAAAG
56851 AGGCCCTGCC CTCTAGAAAG CTCAGATCTT GGCTTCTGTT ACTCATACTC
56901 GGGTGGGCTC CTTAGTCAGA TGCCCTAAAAC ATTTTGCCTA AAGCTCGATG
56951 GGTTCTGGAG GACAGTGTGG CTTGTCACAG GCCTAGAGTC TGAGGGAGGG
57001 GAGTGGGAGT CTCAGCAATC TCTTGGTCTT GGCTTCATGG CAACCACTGC
57051 TCACCCCTCA ACATGCCTGG TTTAGGCAGC AGCTTGGCT GGGAAAGAGGT
57101 GGTGGCAGAG TCTCAAAGCT GAGATGCTGA GAGAGATAGC TCCCTGAGCT
57151 GGGCCATCTG ACTTCTACCT CCCATGTTG CTCTCCAAAC TCATTAGCTC
57201 CTGGGCAGCA TCCTCTGAG CCACATGTG AGGTACTGGA AAACCTCCAT
57251 CTTGGCTCCC AGAGCTCTAG GAACTCTCA TCACAACTAG ATTTGCCTCT
57301 TCTAAGTGTG TATGAGCTTG CACCATATT AATAAATTGG GAATGGGTTT
57351 GGGGTATTAA TGCAATGTGT GGTGGTTGTA TTGGAGCAGG GGGAAATTGAT
57401 AAAGGAGAGT GGTTGCTGTT AATATTATCT TATCTATTGG GTGGTATGTG
57451 AAATATTGTA CATAGACCTG ATGAGTTGTG GGACCAAGATG TCATCTCTGG

57501 TCAGAGTTA CTTGCTATAT AGACTGTACT TATGTGTGAA GTTGCAAGC
57551 TTGCTTTAGG GCTGAGCCCT GGACTCCAG CAGCAGCACA GTTCAGCATT
57601 GTGTGGCTGG TTGTTTCTG GCTGTCCCCA GCAAGTGTAG GAGTGGTGG
57651 CCTGAACCTGG GCCATTGATC AGACTAAATA AATTAAGCAG TTAACATAAC
57701 TGGCAATATG GAGAGTAAA ACATGATTGG CTCAGGGACA TAAATGTAGA
57751 GGGTCTGCTA GCCACCTTCT GGCCTAGCCC ACACAAACTC CCCATAGCAG
57801 AGAGTTTCA TGCACCCAAG TCTAAAACCC TCAAGCAGAC ACCCATCTGC
57851 TCTAGAGAAT ATGTACATCC CACCTGAGGC AGCCCTTCC TTGAGCAGG
57901 TGTGACTGAC TATGACCTT TCCTGGCCTG GCTCTCACAT GCCAGCTGAG
57951 TCATTCCCTA GGAGCCCTAC CTTTCTATCC TCTCTATATG AATACTTCCA
58001 TAGCCTGGGT ATCCTGGCTT GCTTCTCA GTGCTGGGTG CCACCTTGC
58051 AATGGGAAGA AATGAATGCA AGTCACCCCCA CCCCTTGTGT TTCCCTTACAA
58101 GTGCTTGAGA GGAGAAGACC AGTTTCTTCT TGCTTCTGCA TGTGGGGAT
58151 GTCGTAGAAG AGTGACCATT GGGAGGACA ATGCTATCTG GTTAGTGGGG
58201 CCTTGGGCAC AATATAAAATC TGTAACCCCCA AAGGTGTGTTT CTCCCAGGCA
58251 CTCTCAAAGC TTGAAGAATC CAACTTAAGG ACAGAATATG GTTCCCAGAAA
58301 AAAACTGATG ATCTGGAGTA CGCATTGCTG GCAGAACAC AGAGCAATGG
58351 CTGGGCATGG GCAGAGGTCA TCTGGGTGTT CCTGAGGCTG ATAACCTGTG
58401 GCTGAAATCC CTTGCTAAAA GTCCAGGAGA CACTCCTGTT GGTATCTTTT
58451 CTTCTGGAGT CATAGTAGTC ACCTTGAGG GAACTTCCTC AGCCAGGGC
58501 TGCTGCAGGC AGCCCCAGTGA CCCTTCTC TCTGCAGTT TTCCCCCTTT
58551 GGCTGCTGCA GCACACCCCC CGTCACCCAC CACCCAAACCC CTGCCGCACT
58601 CCAGCCTTTA ACAAGGGCTG TCTAGATATT CATTAAACT ACCTCCACCT
58651 TGGAAACAAT TGCTGAAGGG GAGAGGATTT GCAATGACCA ACCACCTTGT
58701 TGGGACGCT GCACACCTGT CTTTCTGCT TCAACCTGAA AGATTCCTGA
58751 TGATGATAAT CTGGACACAG AAGCCGGGCA CGGTGGCTCT AGCCTGTAAT
58801 CTCAGCACTT TGGGAGGCCT CAGCAGGTGG ATCACCTGAG ATCAAGAGTT
58851 TGAGAACAGC CTGACCAACA TGGTGAAACC CCGTCTCTAC TAAAAATACA
58901 AAAATTAGCC AGGTGTGGTG GCACATACCT GTAATCCCAG CTACTCTGGA
58951 GGCTGAGGCA GGAGAATCGC TTGAACCCAC AAGGCAGAGG TTGCAGTGAG
59001 GCGAGATCAT GCCATTGAC TCCAGCCTGT GCAACAAGAG CCAAACCTCCA
59051 TCTCAAAAAA AAAAAA (SEQ ID NO:3)

FEATURES:

Start: 3000
Exon: 3000-3044
Intron: 3045-45393
Exon: 45394-45525
Intron: 45526-45761
Exon: 45762-45818
Intron: 45819-50154
Exon: 50155-50329
Intron: 50330-51076
Exon: 51077-51132
Intron: 51133-52775
Exon: 52776-52933
Intron: 52934-55922
Exon: 55923-56064
Stop: 56065

FIG.3-24

CHROMOSOME MAP POSITION:
Chromosome 22

ALLELIC VARIANTS (SNPs):

DNA

Position	Major	Minor	Domain
941	A	T	Beyond ORF(5')
2612	G	A	Beyond ORF(5')
5080	G	A	Intron
6599	-	A C	Intron
6983	C	G	Intron
9885	A	-	Intron
12538	G	T	Intron
17707	T	C	Intron
18219	-	A	Intron
19670	C	TT	Intron
21153	G	T	Intron
24566	C	-	Intron
26604	G	A	Intron
27255	C	G	Intron
27399	T	C	Intron
28088	G	A	Intron
28734	G	A	Intron
29246	-	T	Intron
29490	G	AC	Intron
29934	T	C	Intron
34480	A	G	Intron
38812	T	C	Intron
40731	C	G	Intron
41303	T	A	Intron
41305	-	AC	Intron
41457	G	C	Intron
43168	A	- T	Intron
43357	T	G	Intron
45664	T	C	Intron
47549	A	C	Intron
47908	C	A	Intron
52267	C	A	Intron
54654	T	C	Intron
54679	C	G	Intron
54693	A	C	Intron
54706	T	C	Intron
54712	T	C	Intron
54799	T	C	Intron
54819	G	A	Intron
55499	C	T	Intron
56825	C	A	Beyond ORF(3')
58871	T	A	Beyond ORF(3')

Context:

FIG.3-25

DNA
Position

941	GAGTAAGTGGGTGGTCAGGTTACAGACTTAATTTGGGTTAAAAGTAAAACAAGAAC AAGGTGTGGCTCTAAAATAATGAGATGTGCTGGGGTGGGCATGGCAGCTCATAAACTG ACCCCTGAAAGCTCTTACATGTAAGAGTTCACAAATATTCCAAAACCTGGAAAGATTCT TTGGATGTGTTGTGTTCTTAAATCTCTACTAATTCTGTTGTCCACTGTCGTAA CCCAACCTGGGATTGGTTGAGTGTGAGCTCTCAGACTTCTGCCTGGAGTTGTGAGAG [A, T] GATGGCATACTCTGTGACCACTGTCACCCCTAAACCAAAAGGCCCTTGTACAAGGGAG TCTGAGGATTAGACCCAGGAAGAATGAGTGTGATGGGCATATATATATCCTATTACTGAG GCATGAGAAGAGTGGAAATGGGTGGTTGAGGTGGTTAAGGCCTTGTCCAGCTGT TTAACTCTCTGGGAACGAGGGGACAACGTGTACATTGGCTGCTCCAGAATGATG TTGAGCAATCTTGAAGTGCCAGGAGCTGTGCTTGTCTATTGATGGCCCTGTGCTGTG
2612	TGAGTTGGAACAGTTGATACCAAAACATCCCCCGCCCCCAACCCCCAGCCTAGGGT CCGTGGAAAAATTGGCCCTGGTGCCTAAAGGTTGAGGACTGCTGATCTAGAGGACCAA TTTATTCAATGTTGGTTGAGTAATGAGCTCTGGATTAGGTGATGGAAAAATCTGAAA AACAGGGCTTTGAGGAATAGGAAAAGGCAGTAACATGTTAACCCAGAGAGAAGTTCT GGCTGTTGGCTGGGAATAGTCATAGGAAGGGCTGACACTGAAAGAAGGAGATTGTGTT [G, A] TTTCTTCTCTCAGAGCTATAACCAAAAGGCTGAAAGTTCTAGAAAAAGGCAAGTTGTT TCAGTAGAAAAAGGATAATCAGAACCTTTAGAAAATGGAATGAGACTACTTTGAG GCCATGAGTTCTGTCCCTGGAGAGATGAGCAGAGGTTGGACAAGTGTCTTACAGAGAT CTTGTGGAGGCAGAAACTGTGCATCTAGCAGAGCATTGGCTAACCTTCAAATGAGAT GCTGTTAACTCAGTCTTACATGGTAGGAATCCTGTCCTTGCCTCTGCTACTT
5080	ACAACGTAAAATAGTTGAAATTGTTGGTGGAAAGAAGAGCAGTCACCTCCAGAGGCTGG ATGGGCATGCCTGGCCCCCAAGGTCTGAAGTGGTAGGGCTGTGCCTATATCCTGAGAATG AGATAGACTAGGCAGGCCACCTGTGCTGTAGATTCCAGCTCTGCACATAGCTCTGTTG TAAAACATCCCTGTGCTTACCAAGTAATTGAGTTGACCTTAAACACTTGCCTTCC CTGGGAACCATATAGGGGATTGGCCTGGAGACGTCTGGCTCTGGAAAGAGTTGGAAAGCA [G, A] CCATCATTATTATCCTTCTTCAGCTATAACTCAGAGCTCTCAAGTCCTTGTGGA TCTTATTGCCTGGTTCTGCCCCCTTACTCCCAGGGAGTTGATTCTGTCTTCTGT TCCATTAGTATGACAGGAGCAGAGAATGTCAAGAGCTGTAAGGGACCTTATAGTTAAAGC CTTGGCTGGCTTCTTCTTACAGTGGGACTAATAAGTAACGTCAAAACCCAAATGAG TTCACAGATTGGTCTGCCTGGCATGTAACCCATATGTTCATATTCTGCTTCC
6599	CTGTAATCCTAGCACTCTGGGAGGCCGAGGCAGAAGGATCGCTGAGCCATGAGCCAG GAGTTGAGACCAGCCTGGCCAACATGGCAAAACTCCACCTCTACAAAAATACAAAAAT ATTAGCCAGGCGTGTGGCACACACCTGTAGTCCAGCTACTTGGGAAGCTGAGGAGCGA TGATTACCTGAGCCCAGGGATATCAAGGCTGTAGTGAGCTGTGATCATGCCACTGTACTC CATCCAGCTGGGGACAGAGTGAACACCCCTGTCTCAAAACAAACAAATGAAAAAA [-, A, C] CCTTAATAATCAGTAACGTCACTTATATTATGTTGTGAGTGTGTCTATATACACCT ATATGTATACATTCTTATTACACATTGGTGTGATCTGATGTGAGCCCCAGGGAT TAAGGGCAACTTGAACATACCCCTGACACAATCAAGCCAATATCATTCCGTGGAGGAAG TAGAGTATCTAGGTTCTGTCTCTAGTTGCAGCTTACCTTGAGGACAGAGACTCTAAC CAGCTGTGCTGAAGGAGCACATCTCTGACTCTGAGCTTCCCTGTAATTCAAAC

FIG.3-26

6983 CACATTCACTGGTATCTGATGTGGAGCCCCAGGGATTAAGGGCAACTTGAACCTACCCCT
 GACACAATCAAGCCAAATATCATTCCCGTGGAGGAAGTAGAGTATCTAGGTTCTGTCTCC
 TAGTTGCAGCTTACCTGAGGACAGAGACTCTAATCCAGCTGTGCTGAAGGAGCACATC
 TCCTGACTTCTGAGCTTCCCTGGTAAATTCAAACGGATGTCACGGCGCCCTCAGATA
 GAGCCTGGTAATTGCCCTGGGAGAGTGACTGTCTTGGATCTAATTGACTTTGCC
 [C, G]
 CAGTTGGAGGAAAATCTCAGGGCTAGGAAGGGATTGTATTTGTCGACCCCCAGAGATAAC
 CTGGGTTTGAGGAACATGGGGCATCAACCTGAATGGCTTGTAAAGATCTCTCCACGCC
 AGCTTGCACTGTTCTGTGATGAAATTAGAGTACCTGAGTAGTGCAAGGCCCTGCTGGAG
 GAGGAGCTCCCTGTGCTACTCAGAGAAATTCACTTCAAGGCCCTCCAGCCTT
 GCTCTTACCCAGCTGGCTACAGTTACAATAAGGAAATGACTTTCTCTCCCTTCCC

9885 GGC GTGCCACCACACCTGCCATT TTTTTTTTTTTAAGTAGAAACAAGGTCTTATTAAAT
 ACTATGTTGCCAGGGCTGGCTTGAACCTCAGCGATCCTCTGCCAGCCTCCCCAAAGT
 GCTTGGGATTACGGAAAGTAAGCCACTGTGCCCTGGCCAGTGCAACCCCCATTTTATCTAA
 AACAGGAAGGCCAGAAAGGTTGGAGTAACTTGTCCAGGGTCACACAGATGATATTGA
 ACTCAGGTCTCCCTGGCTCCCAAGAGAGTCTGCTTCCACTAGGACTCCAGGAGAAAAA
 [A, -]
 AAAAAAAAAAAACAGTAGACTTGGAGACAGAAAATCTGATTGAGTCTTAGTTGAGCTAGG
 CTAACTGTGTAACTGTGGGCAAGTCTCTAGCCCTGTGAGCCTCAGTTCTTATCTGA
 AAATGTCAAAAAGAAATCCATCTCATGGAGTAGTTGTGATGATCAAGGACTCTGAAAAC
 ATTAGAATGGTTAATGTGAAGGGATTAGCAGCAGCACATGGCAACATTGTGCATCTTATA
 TTAACTATCAAATATCAAGCGTCACTGCTATATAAAAGTCATCAAATTAGGCAC

12538 ACTTGGGAGGCTGAGGCAGGAGAATCACTGAACCTGGGAGGCAGAGGTTGCAGTGAGCC
 CAGATCACGCCACTGCACTCCAGCCTGGTACAGAGTAAGACTCCATCTCAAAAAAAA
 AAAAAAAAAAAAATTCTTAATTGGCCTACAGTAGAGCCTCCGTAAATGTGGCCTCT
 CCACATCTCCACAACCTCCTGCTCCCTGCACTTCAGCCTCACCTCTCTGGACAGGCC
 CTCTCTGACAAGGGCTTGTCACTGCTCCCTGCTAGAATGCCCTACTCT
 [G, T]
 TTCACTTAACTCCTGTTATCGTTAGATCTTACCTGGATGGCTCAGAGAAATATAGAA
 GTAATTCTCACCTGAAAAATAGGTTAGGTCCTGTTTATGTTTCTAGACCTTCC
 TTTGAGGCTTTTAAAAAAGTAGTTAATCTCACATTATTCATGTGATCATCTCT
 TAATGATATCTTAAGACCTTAATAGAACAAATTGGTCATGGACTGTGGGTTTGGCC
 CTCATTGTGTCAGCACTGAGCATATTGTTGGCATAGGAGGGATATTGTTGAATGAATTG

17707 GTAGTGGGTGCTCAGAGTGTGTTGGTGAATGATGTATTGTTGAACGACTTTGGA
 CACTGAAATAAAAGTCATCCAGTATGCCATTACCATCTTCGCTCTACAATATTCTT
 TTAGGCAAGAGCTTATCTTGTGGTGAAGATAAGCTCAAACCTATGTAGACTAAC
 CTCAGTCTGAAATGTCACTCCCTAAGTCTTAAACCATCAAAACGCCAGGGCTCAAGGAATG
 GCATGCCCTCTGCAACTGTAGCAACCTGCTGTGTTATTGCGTGTGTTCTGATT
 [T, C]
 CCCCCAAAGCTAGAGTCCCTCTCCATGGCAGTGCTGGAAAGTGTGCTAACAAATTCTT
 CTCCATACTGCTTACGATTACAAAAAAACCCCTCAGCATCTCATGCCAGACTTGAGTTAA
 GGTTGTTCTTTGTGTCAGCTGTATTCTGGTCATGACTTCCTGATGATGCCCTATA
 GAGATTTGCTGAGATCAGAGGGTGTCCACTGCCATCAGTAGCACTGACTTTGAGAA
 GCACCGTTCTGAAGTGGCTAATGTCACTCCCTACGTTGTTGAAATTGTTT

FIG. 3-27

18219 TGCCATCAGTAGCACTGACTCTTGAGAACGACCGTTCTGAAGTGGCTAATGTCATCC
 CTCACGTTGTTGAAATTGTTAGTCCAGAGATAGCCTTCAATGGAATGAC
 GCTATCTCTAGAATCACTTTTTTTGAGTTGGAGTCTCGCTGTGTCGCCAGG
 CTGGAGTGCAGTGGCACAATCTCAGCTACTGCAATCTCCACCTCCGGGTTCAAGTGAT
 TCCCCTGCCTCAGCCTCCCAGGGAGCTTACTACAGGCGCACACCCCCACTCCTGGCTA
 [-, A]
 TTTTATGTGTTTAGTAGAGACGGGGTTTACCGTGTGGCCAGGATGGTCTCGATCTCC
 TGACTTTGTGATCTGCCTGCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGTCAC
 CGCGCCTGCCCTAGAATCACCTTTTATACCATAACGTGAGCACCCTGCCGCGTCACCA
 AGGAAAGAGAGAGGGCAGCTACTGTGGGTTACAAATGGTAAGAGTGGCACCAGGAAGGT
 GAAAGTCTACTTAGCCAAGGCTAACAAATGTCAATCACCAAACATTATTAA

19670 GACCCCCATGATGAGCAACTATAGCACTAGAACAGTGATAATAACTAATGTTATAATGC
 ATCTTCAGTTACAGAGGGTTTGTACTCATCATCTAGTTAGTCCCTGCAACAAACCTC
 TTGAGGAATATAGCACAAGCAGGACAAGGGAAAGCCCAGAGATGTTAAATAATTATCAA
 GTTTATGCTGCTGGGAAGGGCAGCACTGAAATTAAAGAAAAGTTTCTGAGCTCAAATC
 CCATGCCCTTCCTCAATGTGAGCTAGCAAGGTATTAGGAATCTGCCTACAGTT
 [C, T]
 AGAGCCTCAAATTGCTGGGTATGTTGAGTTCTGTATCTGATTCTAGATTTCTGCC
 CACATTCTACTGTCTGGATATCAGGAAAGAGTTATCAAATGCCGTGGAAATCCAAGA
 TAAGGTCTCATGATGAGTAACCCAGTGAAACATGAAGTCAAGTCTAACTAGTCACT
 ATTTCACTACTGCTGACTCCTGATGATCAGCTCCTTCTAAGTGCTTACTGTCCACTTA
 TTCCATCATCTGCCCTAGAATTATGTGAAGGAATCAAAGCAAAAGGATCATAAGGCTCC

21153 GGACCCCTGTTAGAAGGATGACTGCTGCTATAATGTAGAAAGTGATTGGAAAGGGGG
 AGGAGTGGGGCACGAAAGATGGTAGTAGATGGGGTGGTAATGCTTACCTTCAGTATT
 TGGAGGCTCGGAGTCCTCAAAAATTCTTCTGATTGGAGTCTCCAGCCAATAGA
 GGGCTTCACACAAACAGTTCTGGTTGAATTGTTGACCAGAGCTTCTCCGACA
 AAAGGTTGGGTGATTCACTTACACACCTGCCCTGAACATTCACTTGGGCTGCC
 [G, T]
 GTTATGAAGGCTATTGTTCTCCAGCCTGTCACAGACGTTGAAGACCTGTGCCCTAGCT
 GGTTCTAAGGAGTCAGTTGTTCACTCCGTGCCAGGTTCCAACCTATGAAATGTGCTG
 GAGATTAACACCTCTCTGCCATTATCCCTACTATAATTGCCAGTCAGGATTCTG
 CAGTTGCCCTCGCAGCCATAACTGATGAATGTTCTGCCAGCTGCTTGAGGACCTAGAA
 GAGCAGTTTCTATCCAGGACCAGTTCCAAGGGTGGGAGGGTGAATATATCCTCCAGT

24566 CTACTCTGGAGGCTGAGGTGAGAGGATCACTTGAGTCCAGAAGGTGAGGTCAAGATTGT
 AGTGGCCATGATGGCATCCGCACCTCCAGCCTGAGTGACAGAGAGAGACCCCTGACTCA
 AAAAAAAAAAAACAAAAAAACACCCCTCACCACCTATCAGCTATTGTCTTGAGAA
 TAGTGACATAACCCCTCAGAACCTATTCTCTAAATCTGTTAAATGAGGCTGATGACGTT
 CTCCCTTACTGGCAATTAAACATGATGGATAATAATGCTAACACACAGGGC
 [C, -]
 TAGAAGATATTAACTGCTCAATAATGGTAGCTTAAACAGTATTCAAACCCATGTGCT
 CTTATCACATGCATTGTTGCCCTGTGTCAGTTGGTGGAAATGGAAAAGGCTCCCTTG
 AACCCCATCTACCATCTTATCAGACTTCTGCCATGGTCAGTAAGAGATAGAAGC
 TGACGGTGAATTCTGGCTCTTACAATGGTAGCGGGTGTGTCCTGGTAAGGGAGAGCT
 GATGTCACTGCCCAAATCCAGTAGTGAGATCTGAGTGTGTTCTGGTTCTCCAGCAGCCT

26604 GATTTGCAGCTGAGCTGTCTATCTGGTGTGGGAAGAAGATGGGGAGTTACTTGTCAAGTC
CCGGCTTACTTCACCTCCAGAGACCTGTTGGTGAAGTTGGTCTCCAGTTCCCTCTCC
ATCTCTCCTGGCCCCCTGGTCTGAGAGGGAGGGTGGTCTCCCTAAATCTCCTTCTCACTTA
GTCCCTTACCATCGGTTCTGCCGGGAGAAGCCAGCGGAGGTTACCCAAGGAGAATCG
GCCTTGTGAGGTACCCCCATTATGTCTGGAAAGTGGTGAAGGGAGGGATATACCCAGAAG
[G, A]
AACTTCTTAGGGAGCTCCAGCTCCCTCTATCCCAGACAAACCTGAAGGAGCCTCCAAA
AGATGCCACTGACCTGCCATTGTAGATGTTACTGCTCCGGGGGAATAGCCCCAAATAG
AGTGCTTCTCAGCTCACATGTCTTACCTGCCGGGAGTGGCTGCTGCCAGGAATT
GTCCCAACAAGCAGGATGGCAGGTTTGCCAAACTGTGGAAACTGGCAAGTCTGGGTG
TGGTAGCTGGTACACAGTAGGCACCTATAAACGTTTGTCTTAATGGCAGGCACA
27255 TGGGGAAAGACCTGGCGAGTGCTTCTAAGACTGGAGCAATGGCTTAGAGTGTCTG
AGCTGCTGGCCAGCCCCACACCTCCTCAGTCCCTAGGCCAAGTACCTCACAGAGCCT
CTCTCTGTGGGGCTTCTCAGAGGGAGATGTGGAAACTCTACCTCTAACCTGGCTTCTT
GCTCATTGCCCACTCCACCTCCATAGAAACTCCCCAGGGGGTTCTGGCCCTCTGGGT
CCCTCTGAATGGAGCCATTCCAGGCTAGGGTGGGTTGTTCTTCAATTCTGGGAGCAG
[C, G]
CTGTTGTTCCAAAAAGGCTGCCCTCCCCCTACCACTGGTCTGGTCGACTTTCCCTCT
GGCTCTCTAAGCTAGGTCAGTGCCAGATCTGCTGCCGGATACTAGTCAGGTGGCC
AGGCCCTGGCAGAAAAGCAGTGTACCATGTGGTTTGTTGAATGACCGGACCCCTGGTAG
ATTGCTGGAAAGTGTCTGGACAGGGGGAGGGGAAGGGAACTGGTCTCAATGCTGACT
CTACCAAGGCCCTGCTAGACACTTATCCTTAATCTCTAACAGCCTAAAGAGATTAT
27399 AGATGTGGAAACTCTACCTCTAACCTGGCTTCTTGTCTATTGCCCACTCCACCTCCC
ATAGAAACTCCCCAGGGGGTTCTGGCCCTCTGGGTCCTCTGAATGGAGCCATTCCAG
GCTAGGGTGGGTTGTTCTTCAATTCTTGGGAGCAGCCTGTTCTCCAAAAAGGCTGCC
CCCCCTCACCACTGGTCTGGTCGACTTTCCCTCTGGTCTCTAAGCTAGGTCCAGT
GCCAGATCTGCTGCCGGATACTAGTCAGGTGGCAGGCCCTGGCAGAAAAGCAGTG
[T, C]
ACCATGTGGTTGTGGAATGACCGGACCCCTGGTAGATTGCTGGGAAGTGTCTGGACAGG
GGGAAGGGGAAGGGAACTGGTCTCAATGCTGACTCTACCAAGGCCCTGCTAGACACT
TTATCCTTAATCTCTAACAGCTAAAGAGATTATATATCCCCATTACAGATGAGGC
AACCAGTTCAACAGAGTTAACATATGGAGCCTACTGGCAGCTTCTGTCTCCTG
ACTTCTCTCATCCTCAGGGGGCTGCAGGTTGTTCTCTCTAGTGGAGAGGAAT
28088 AAGAGCCAATGGAAATTGATCTTGAGTTAGGAGAAAGCTTACATGTGGATTAAAGAT
GCCAAGTGTGAAGTAGCCACATTCAAGGCTCTATTAAATTCTCTTAATCCTGGGAAGG
CAGCTAGGAGAAGGGTTCTCCCTTAGGAGCCAGGAACATACCCCTTACCTTGGGA
GAGGCAGGGAGCCAGGGAGGACACAACCTCTCAGGAAGAGGAGAAGCTAGAGCAGATAG
TGAACCTCAACCTGAACCTTAAGGGCCAGACCAACTAATGCCACCCAGTCCACCTGCC
[G, A]
TTTGTCTTGTCTGTCCAGGCTTCTGGAGAACCTGATCTCTGCCCTACCCCCAAG
CTCCGTTGCCAGCTAGAGTCTGGGGGGTACTGACTGACTTCTGAGACATTCTCCCT
TCCCCAAATAAGAGGCCACATTCTGAAGTCACCTCTGAAGAGATAGCTGCCACACAGGG
CTCTTCCCCCAGGGAGGGACCACCCAGACCCCTCTGCTCTCCAGGTATCGTTACAC
ATCACTACCTGGTCAGAAAGCTGTTCTGCCATTAGCCCTCCCTTTTATTATAGGAT

28734 AAGTAGAAGCTAGACTTCTGGCTCCTGAACAGGGCCTTGCTGGATTCTGTGAAACAA
 ATTAAGTCTTGACCTAGGCCTCTGGGGAGTACAAAGTCTATGGAGTTCTGGGCTG
 TGGTTGCAAGGAAAGTGACGCAACCAGATTCCATGGGACATGATCAGGCGTACATGTG
 AGGGAGGAAGAGGGAGCAAGGGAATGAAGAATACAACCTCTGTGTCCTACACCCCCTG
 CTGACAGGCCATACATACTCAGCAGAGAATGCACTGTCTTCTACCACACTAGCGTGAG
 [G, A]
 AGTGAGCTGCAATTACCACTGTGCTTCAAGTAAGAAAATACCTCAAATTGGAATTTACA
 AAAGAGGTAATTAGGGAGTGGCTTTGTCGGACATTTAAAGCATTTTCTTTTATA
 GAATTCACTTAATGTCACACTGATTTAATGAGCTGGTTACACATTATCTCTG
 AGAAAACAAATGAACCTTGTGTTCAAAGCAATCCATGTTAAAGGGAAAAATTATGC
 ATAACCTGCCCAGCTCACAGTAACCTTGGCAGGTGCCTAGGTCTGGGACTCT

29246 AATCCATTTAAAGGGAAAAAAATTATGCATAACTCTGCCAGCCTCACAGTAACCTTGC
 GCAGGTGCCCTAGGTCTCTGGGACTCTTCTTATCTGAAAAATGAAGGACTTGGATC
 AGGTGAATGGTCCAGCTCTGCAACTTATGTGGCTCTCAGAGGACACAAGCTCTT
 CCATTATTGCCAATAATGGAGGCCCTGCTTTACTGCACTACAACACACAAATAC
 TTGAAAATACAGTCTCTGGTTTTGGTTGGAACTGAATCAGTGCACCTAGCAACACT
 [-, T]
 ATTTCTGCTGTTGCTAGGCTTCATTATGTGTTGGTTAATTAAAACAACAATAAC
 ATATTCCATAATAATTACAGCTTAATTGGCAGACTGTTCACTAGTCTAGGATCTGCAGGA
 AGGAGGAGTAATAAAGGGATTTTGACTGAGCTCTTATGGAACAGAGTCTCTAGGCC
 CTGTCATATCTGCCCTCTGGGCCCTGGGAAAAGTTGGCATCCCCAGTTGTGGTGTCT
 CCAGGTGCCCTCAGGCTGTGGTGGAGGGAGCTCCATTCTCCTTCAGCCCACACTCAAT

29490 AACTACAGTCTCTGGTTTTGGTTGGAACTGAATCAGTGCACCTAGCAACACTTATT
 TCTTGCTGTTGCTAGGCTTCATTATGTGTTGGTTAATTAAAACAACAATAACATA
 TTCCATAATAATTACAGCTTAATTGGCAGACTGTTCACTAGTCTAGGATCTGCAGGAAGG
 AGGAGTAATAAAGGGATTTTGACTGAGCTCTTATGGAACAGAGTCTCTAGGCCCTG
 TCATATCTGCCCTCTGGGCCCTGGGAAAAGTTGGCATCCCCAGTTGTGGTGTCTCCA
 [G, A]
 GTGCCCTCAGGCTGGTGGAGGGAGCTCCATTCTCCTTCAGCCCACCTCAATTCA
 AGGCTAGGGGCTGAAAGAAGCTCTCTACAACACTGGCTGTTCACTGGAGGTTAAGGGATG
 ACCATCCAGCCAGGCCCTCAGGACATGGGAGGGCTTATGTTAACATGTGTAACATT
 CACTGCAATAATGACTGGTTCTTACCCATAAGGTTGAGAATTACCTGTAACATT
 TTGTCAGGAAATTGGATGTAAGTGAGGGCTGGCCTATCTACTTCACTGGCTTC

29934 GGACATGGGAGGGCTTATGTTAACATGTGAAATCCACTGCAATAATGACTGGTTCTT
 TTACCCATAAGGTTGAGAATTACCTGAAACATTGGTCTGAAAGAATTGGATGTA
 GTGAGGGCTGGCCTATCTTACCTCACTGGCTCTCAGCACAGCACCTGGCTGC
 TTGTTCTACACATCCTAGATGCACAGTAACATTCTAACATTAGAAATCTATTAGA
 ATCAATTGATTCTAGCTGGCTGGCTCTGTAATCCAGCACTTGGGAGGC
 [T, C]
 AAGGCTGGAGGATCACCTGAGTCAGGAGTTAACGACAGCCTGGCAACATAGGGAGAC
 CCTGTCCTACAAAAAATTTAGCCAGGATGGTGGTGTGCACCTGTAGTCCCAG
 CTACTCAGGAGGCTGAGGCAGGAGGATCTCTTGAGCCTGGGAGGTGAGACTACAGTGAGC
 AATGATTGTGCCACTGCACTCCAGCCTGGGTGACAGAGTAAGACTCTGTCTTAAAAAA
 AAAAAAAAGTTGATTCTATTGGATAGATAAATAATTCTTCTAGGACCTTCTT

34480 CTGACTTCAAGTGATCCACCCGCCTCGGCCTCCAAAGTGCTGGGATTATAAGCATAAGC
 CACTGTGCCAGCTGCTCTATATTTTATACATATTATTCATTAATTTCACAGC
 AGTTCATTTATAGATGAGGAAACTAGGCCAGAGAAGTAAATATCTTGCCCAAGATGAT
 GTAACTAGTAAGTGGCAGGATCAAGATTAAACCAAGCAATGTTCAAACCTTGGAAAGC
 AAGAATGTGGCCACTGTGGAGGTCAAGGCCCTGACAACAAGAATAGGGAAAAGAAGGA
 [A, G]
 CTAGAAGGAAAGAGATGGCATGGCTCAGCAGGCCAGGGAGCTTAGCTGTGTGTTG
 GGAAGCTCAGAAGGGAGGAAGAGGTTGTCAGCAGGTAAGTCTGAGAACACACCAGAC
 TTTGAGAGGTGGAGGCTCATGCCAGGTATTAGGGGAGAAGGGAGCTATAGATTTTT
 TTTTTTTTTTTTTTTTTTTTTTTTTAGAGACGGGCTTACTATGTTGCCAGGCTG
 GTCTTGAACCTCTGGCTCAAGTGATCCTCCCACCTCAGCCTCCAAAGTGCTGGGATTA

38812 AAATCCACCGAGATCCATTGAGAGTTAACGCAGCAAGGTGTTGACCAAGTTAACATTT
 AGAAGGATCACTGGTATGGAGGTTGGATTGGAGAGGGAAAGCCTAAAGGTATAGAGACT
 AGTTAGGAAGCTATTGAGGCTGGGATGGTGTAGGCTGGGTTCATGCCCTGTAATCTCAGCACTTGG
 GAGGCTGAGGTGGGAGGATTGCTGAGGCCAGGAGTTGAAGACCAACCTGGCCAACATAG
 CAAGACCCGTCTGTGTTCTTAATTAAAAGAAAAGTCCAGACGTAGACATAGTGGCT
 [T, C]
 ACGCCTGTAATGCCAGCACTTGGGAGGCCAAGGTGGGAGATTGCTTGAGGTCAAGAGT
 TTGGGATTAGGCCAGGCCAGTGGCTACGCCCTGTAATCCCAGCACTTGGGAGGCCAG
 GTGGGCGATCACAAGGTCAAGGAGATCAAGACCATCCTGGCTAACACAATGAAACCCCGT
 CTCTACTAAAAGTACAAAATTAGCCGGGATGGTGGGGACGCCCTGTAAGTCCAGCTAC
 TCGGGAGGCTGAGGCAGGAGAATGGCTGAACCTAGGAGGCCAGCTTGCTGTGAGCAGA

40731 GTTCTGCTCATGTCTGCTCTGGATGAAGCTGAGCTGGCTTCAGAAGCCTGCAGAGT
 TAGGAAAGGAACCAGCTGGCCAGGGACAGACTATGAGGATTGTGCTGACCCAGCTGCC
 TGTGGGATCACAGTTACAGCCAGAGCCTGCGGACCCAGCTGCTGCCAGGTTCT
 TAGAAACCTGAGAGTCAGTCTGTCCACTGAACTCTAACGCTGGACAGGAGGCACTGAT
 GCTAAACCTGAAGGGCAACATGCCATGGAGAAAGCATGGAGCTAGAGCCTGGAGTA
 [C, G]
 GGGCACAGATAGGATTGAATAATTGTGAGAAAGACTTTGAAAACAATAAGCAAAAGA
 TGAATGAACGTTTTTTAGACTTGAGGGACCAACAACCCCAACCCAGATTCTGCCA
 GGTCCATGGGAAAGGAGAAGTTGCCCTGAGTGGAGGCCAAGTAGGGAGACTTACAGAA
 AAGAAGTCAGAGCACTGGCTCCAGGCAGAAATACTGATACCCACTGGGCTTCAGGC
 TGAGCTCCCTTACAAATCACTCTGAGCCTGTTCTGCATCTGTGACAT

41303 CTCTGAGCCTGTTCTGCATCTGTGACATAAGATGGTAAGATAAAGGTGGCTGTCACC
 AATTATGTAAGGATTAATGTGAAAGGACATAAGTTGATAGTGCTGCCATAGGGAC
 AGTGTTCAGTAAACGTGACACATTCTAGTATCACTAAGAATCAGGTTCTGGCCAGGCA
 CCGTGGCTCATGCCCTGTAATCCCACACTCTGGGAGGCCAGGTGCGGAGGATGGCTTGA
 AACAGGAGTTGAGACCAGCCTGAGCAACATAGTGAAGACACTGTCTTACAAAAAAA
 [T, A]
 AATAATAATAATTGTTTTAATTAGATGGCAGGGCACTGTGGCTCACACCTGTAATCCC
 AGCACTTGGGAGGCCAAGGCCAGGATTGCTTGAGGCCAGGAGTTCAAGGAGCAGCCTG
 GGCCACATTCTGTCTACAAAGAATAAAAAGTTAAGTGGCATGGTGGCACATGCC
 GTAATCCCAGCTACTCAAGAGGCTGAGGAGGAGGATTGCCAGGCCAGGAGTTCAAGAC
 TGCACTGAGCCTGATCACACCAGTGTACTACAGCTGGCAACAGAGTGAGACCTTGTC

41305 CTGAGCCTGTTCTGCATCTGTGACATAAGATGGTAAGATAAAGGTGGCTGTCACCAA
 TTATGTAAGGATTAATGTGGAAAAGGACATAAAAGTTATAGTGTGCCATAGGGACAG
 TGTTCACTAAACGTGACACATTCTAGTATCACTAAAGAATCAGGTTCTGGCCAGGCACC
 GTGGCTCATGCCTGTAATCCCAACACTCTGGGAGGCCTAGGTCGGAGGATGGCTTGAACA
 CAGGAGTTGAGACCAGCCTGAGCAACATAGTGAGACACTGTCTCTACAAAAAAAATA
 [-, A]
 TAATAATAATTGTTTTAATTAGATGGGCAGGGCACTGTGGCTCACACCTGTAATCCAG
 CACTTGGGAGGCCAAGGCCGGAGGATTGCTTGGGCCAGGAGTTCAGGAGCAGCCTGGG
 CCACATTCTGTCTCACAAAGAATAAAAAGTTAACTGGCATGGTGGCACATGCCTGT
 AATCCCAGCTACTCAAGAGGCTGAGGAGGAGGATTGCTTGAGGCCAGGAGTTCAAGACTG
 CAGTGAGCCTGATCACACCACTGTACTACAGCTGGCAACAGAGTGAGACCTGTCTC

41457 CTAAGAACATCAGGTTCTGGCCAGGCACCGTGGCTCATGCCTGTAATCCCAACACTCTGGG
 AGGCCTAGGTCGGAGGATGGCTTGAACACAGGAGTTGAGACCAGCCTGAGCAACATAGT
 GAGACACTGTCTCACAAAAAAAATAATAATAATTGTTTTAATTAGATGGCAG
 GGCACTGTGGCTCACACCTGTAATCCAGCACTTGGGAGGCCAAGGCCGGAGGATTGCT
 TGAGGCCAGGAGTTCAAGGAGCAGCCTGGGCCACATTCTGTCTCACAAAGAATAAAA
 [G, C]
 TTAACGGCATGGTGGCACATGCCTGTAATCCAGCTACTCAAGAGGCTGAGGAGGAGG
 ATTGCTGAGCCCAGGAGTTCAAGACTGCACTGAGCCTTGTCTCCAAAAAAAAGTTGTTTTTATCCACT
 CTCCTCACAAACAAACTGAGTAAGTTAGAGCCTCTCAGCTGGCATGTGTTGAAACAG
 TGCCCTCTCATTAAAGTGCTGCCCTCACTCCATTGCCTTGGCCTGGTCAGTATGAT

43168 AGCTACTTGGGAGGCTGAGGCAGGAGAACGCTTGAACCTGGAAGGCCGGAGGTCGAGTG
 AGCCGAGATCGTGCCTTGCACCTCAGCCTGGCGACAGAGCAGACTCTGTCTCAAAA
 TAATAATAATAACAATAACTAGCCGGCCTGGTGGCACATGCCTGTAGTCCCAGTTACTC
 AGGAGGCAGGAGCATGAGACTCAGGTGAACTAGGGAGACAGAGGTTGCAGTGAGCCAAGA
 TCACACCACTGCACTCCAGCCTGGTTGACAGAGCAGACTCTGTCTCAAAAAAAA
 [A, -, T]
 CCCATTGCTCATTGGATACTAGTATAACTACTCTAAACCACTGTTAGTACTTAA
 ATCAAGCAGATATGGGAGATGGTGAATTACCATCTACAGTGTGTATATGTCACATA
 CTGAGCATTATCAGCTAGTAACTAGTTAATTGTTATGTGTGATGTATGAGAGTT
 CCCATTGTAATGTGTTTACTATGCTTAAATAATGACTGATGTCAGCAACCCAAAA
 TGATACATCTGATGTAAGAGGCCCTGTTCCCAATAAAACATCTAAACTATAGACATTG
 AGC

43357 AGGCATGAGACTCAGGTGAACTAGGGAGACAGAGGTTGCAGTGAGCCAAGATCACACCA
 TGCACCTCCAGCCTGGTTGACAGAGCAGACTCTGTCTCAAAAAAAAATCCATTG
 CTCATTGGATACTAGTATAACTATCAGCTAAACCACTGTTAGTACTTAAATCAAGCA
 GATATGGGAGATGGTGAATTACCATCTACAGTGTGTATATGTCACATACTGAGCAT
 TACAGCTAGTAACTAGTTAATTGTTATGTGTGATGTATGAGAGTTCCATT
 [T, G]
 AATGTGTTTACTATGCTTAAATAATGACTGATGTCAGCAACCCAAAAATGATACTC
 TGATGTAAGAGGCCCTGTTCCCAATAAAACATCTAAACTATAGACATTGGAATGAACA
 GGTGCCCTAAGTTCCCTCCAGGGTTCTGGCCGGTCTGTGAGGACTACACATCC
 CTACTCCCGTCTTCCATCTCAGGCGCAGTAACAGTATCTCCAAGTCCCCGGCC
 AGCTCCCAAAAGGAGGCCCTGCTGTTAGCCGTGACATCAGCCGCTCAGAATCCCTCGT

FIG.3-32

45664 CCAGCTTCCCTGGCTCCCCACCCCCAGGTGAAAGTGTGCGCAGCCTGGACCAACCC
 AATGTGCTCAAGTTCACTGGTGTGCTGTACAAGGATAAGAAGCTGAACCTGCTGACAGAG
 TACATTGAGGGGGGCACACTGAAGGACTTTCTGCGAGTATGGTGAGCACACCACCCAT
 AGTCTCCAGGAGGCCTGGTGGGTTGTCAGACACCTATGCTATCACTACCCCTAGGAGCTTA
 AAGGGCAGAGGGGCCCCCTGCTTGGCTCCAAAGGACCATGCTGGGTGGACTGAGCATACA
 [T.C]
 AGGGAGGCTCACTGGGAGACCACATTGACCCATGGGGCCTGGACCACGAGTGGGACAGG
 GCTCAACAGCCTCTGAAAATCATCCCCATTCTGAGGATCCGTTCCCCTGGCAGCAGAA
 GGTAGGTTGCCAAGGAATGCCCTCCGGATGGTAGTCCCACCAACAAACCTGCCAG
 CAGGGCGAGAGTAGGGAGAGGTGTGAGAATTGTTGGGCTTCACTGGAAAGGTAGAGACCCCT
 TCCTATGCAACTTGTGTGGCTGGGTCACTGGAGCTATTGAGTTGTCTGTCACTG

 47549 AATTAGCTGGCGTGGTGGTGCACGCCGTAGTCCCAGCTACTCAGGAGGCCGAGGCAGG
 AGAATAGCTGAACCTGGGAGGCAGAAGTTGAGCTGAGCCAAAGATCACACCACTGCATT
 CAGCCTGGGTGACAGAGTGAGACTTCATCTCAAAAAAAAAAAAGAGAGACTGATATG
 GTTAGTACATTGGGTGGAATGGGAGGGTCAGGGATGGAGCCCTGCATAGGGGCTA
 ATGAAACATTCAGATTCTGAATTAGGTAGTGGCTGTGGGACAGGAGCCTGGAGGC
 [A.C]
 GGGTGGAGTCAGAATGGAGAGACTGGTGGCAATGAGGGAACAGGAGGAGGAGGAGGAGG
 AGTTACGAGTGGCTTGAGGTGTCACTTACAGACATTGGGGATGGGGATAGCCGTGA
 TTGTTGAGCAACTGGTTGGGAAGAGCTAGCATTGATCCCTGCTGTTCTGTGCTAGCAGA
 ACCTATCAGCATCTCTGGCAGGAAACTGGCTCCATGAGACTGGCTTAGGGAGAGGCTG
 CTAGTCACCTAATCTGAGAGAAGGGCAGCTGGAGCTGTGGGACAGAAGAGGCATCC

 47908 GGAGTTACGAGTGGCTTGAGGTGTCACTTACAGACATTGGGGATGGGGATAGCCGT
 GATTGTTGAGCAACTGGTTGGGAAGAGACTAGCATTGATCCCTGCTGTTCTGTGCTAGCA
 GAACCTATCAGCATCTCTGGCAGGAAACTGGCTCCATGAGACTGGCTTAGGGAGAGGC
 TGCTAGTCACCTAATCTGAGAGAAGGGCAGCTGGAGCTGTGGGACAGAAGAGGCATCC
 ATGTAAGCTGGTGGGGGTCTCAGCTTGTGAAGAGGAGATGGCTTGTGAGCAGGGCTGACA
 [C.A]
 TGAAAAGGCTGGAAGAAAAACAGACACACAAGAGTCAGGATCAGGTAGCATAGGAA
 AGTTGTTGAGCACTGGTTGGGAGAGCTCCCTCAGGCAGGCAGGCAGGCAGGTATGAGCT
 ATAGCGATTCAAGGAAGAGCTCCCTGGGTGTGAGCAGCTCCAGGAGCCTAAGGGATGAA
 AGTAGTATTGAGGGCTGGAGAGCAAGGAGTGGCTTCTACATTGCAAGGGAAAGG
 AGAAAGGAAGTTGCTCTGAGAGTGGTAAGAGTCAGTGGTGGAGGCCTGGAGAGGAGACA

 52267 TTGTGAGGGTAGAGGGAGAGGACAAGGGATGGTAGGATAATGAAGGAATGTTTG
 TTTTGTGTTTGAGATGGAGTTCACTCTGTCACCCAGGCTGGAGTGCAGAGGT
 GCAATCTGGCTCACTGCAGCCTCGCCTCCAGGTTCAAGCAATCCTCTGCCTCAGCC
 TCCCAAGTAGCTGGACTACAGGTGTGCGCACCACGCCTGGCTAATTGGTATTGCA
 GTAGAGACAGGGTTTCGCCATTGGCAGGCTGGTCTCAAATGCCACCTCAGGTGAT
 [C.A]
 CACCCGCTTCAGCCTCCAAAGTGTGAGATTACAGGCATGAGCTACCGTGCCTGGCCAT
 GAAGGAAGATTGTTAAAAAATTGTTCTTAATTTAATTGAAACACCTCTGTTCA
 AGCACTGGCTGGTGCAGAGGGTTCAAGACATGAATCAGATCCAGCACCTCATAGAGCC
 TTAATCTGGCACACACACAGCACAAGGAGACACAGACAAGGCAGGGTAGGATGAGTG
 GAAGCTAGGAGCAGATGCTATTGGAACACTGGCTCTGAGTGAAGCCCTCTTAG

FIG.3-33

54654 GGCCCCGGCCCCGGCCCCCAGGCCAGGCAGTCGGCGGCCAAGGACCACGCATCTACTTCA
 GAGCCCCCCCCGGGGCGCAGGAGAGGGCCGGCTGGCGGATGATGAGGGCCAGTGA
 GGCACAAGGAAGGTACCATCAAGTATGACCCCAAGGAGCTACGGAAGCACCTAAC
 TAGAGGAGTGGATCCTGGAGCAGCTACGCGCTCTACGACTGCCAGGAAGAGGAGATCT
 CAGAACTAGAGATTGACGTGGATGAGCTCCTGGACATGGAGAGTGACGATGCCCTGGGCTT
 [T, C]
 CAGGGTCAAGGAGCTGCTGGTTGACTGTTACAAACCCACAGAGGCCTTCATCTCTGGCCT
 GCTGGACAAGATCCGGGCCATGAGAACGCTGAGCACACCCCCAGAAGAAGTGAGGGTCCCC
 GACCCAGGCAGCGAACGGTGGCTCCCATAGGACAATCGTACCCCCCGACCTCGTAGAACAG
 CAATACCGGGGGACCCCTGGCCAGGCCTGGTCCATGAGCAGGGCTCCTCGTCCCCCTG
 GCCCAGGGGTCTCTCCCTGCCCTCAGTTTCACTTTGGATTTTTATTGTTAT

54679 GGCAGTGGCGGCCAAGGACCACGCATCTACTTCAGAGCCCCCCCCGGGGCCGAGGAGA
 GGGCCCGGGCTGGCGGATGATGAGGGCCAGTGAGGCAGCAAGGAAGGTACCATCAA
 GTATGACCCCAAGGAGCTACGGAAGCACCTAACCTAGAGGAGTGGATCCTGGAGCAGCT
 CACGCGCTCTACGACTGCCAGGAAGAGGAGATCTCAGAACTAGAGATTGACGTGGATGA
 GCTCCTGGACATGGAGAGTGACGATGCCCTGGCTTCCAGGGTCAAGGAGCTGCTGGTGA
 [C, G]
 TGTTACAAACCCACAGAGGCCTTCATCTCTGGCTGCTGGACAAGATCCGGGCATGCAG
 AAGCTGAGCACACCCCCAGAAGAAGTGAGGGTCCCCGACCCAGGCGAACGGTGGCTCCCAT
 AGGACAATCGTACCCCCCGACCTCGTAGCAACAGCAATACCGGGGGACCCCTGGCCAG
 GCCTGGTCCATGAGCAGGGCTCCTCGTGGCCCTGGCCCAGGGTCTCTTCCCTGGCCC
 CTCAGTTTCACTTTGGATTTTTATTGTTATTAACATGGACTTGTGT

54693 AGGACCACGCATCTACTTCAGAGCCCCCCCCGGGGCCGAGGAGAGGGCCGGCTGGG
 CGGATGATGAGGGCCAGTGAGGCAGCAAGGAAGGTACCATCAAGTATGACCCCAAGG
 AGCTACGGAAGCACCTAACCTAGAGGAGTGGATCCTGGAGCAGCTACGCGCTCTACG
 ACTGCCAGGAAGAGGAGATCTCAGAACTAGAGATTGACGTGGATGAGCTCTGGACATGG
 AGAGTGACGATGCCCTGGCTTCCAGGGTCAAGGAGCTGCTGGTTGACTGTTACAAACCC
 [A, C]
 AGAGGCCTCATCTCTGGCTGCTGGACAAGATCCGGGCATGCAGAACGCTGAGCACACC
 CCAGAAGAAGTGAGGGTCCCCGACCCAGGCGAACGGTGGCTCCCATAGGACAATCGTAC
 CCCCCGACCTCGTAGCAACAGCAATACCGGGGGACCCCTGGCCAGGCCTGGTCCATGA
 GCAGGGCTCTCGTGGCCCTGGCCCAGGGTCTCTTCCCTGGCCCCCTCAGTTTCACT
 TTGGATTTTTATTGTTATTAACATGGACTTGTGT

54706 TACTTCAGAGCCCCCCCCGGGGCCGAGGAGAGGGCCGGCTGGCGGATGATGAGGG
 CCCAGTGAGGCAGCAAGGGAAGGTACCATCAAGTATGACCCCAAGGAGCTACGGAAGCA
 CCTCAACCTAGAGGAGTGGATCCTGGAGCAGCTACGCGCTCTACGACTGCCAGGAAGA
 GGAGATCTCAGAACTAGAGATTGACGTGGATGAGCTCCTGGACATGGAGAGTGACGATGC
 CTGGCTTCCAGGGTCAAGGAGCTGCTGGTTGACTGTTACAAACCCACAGAGGCCTCAT
 [T, C]
 TCTGGCTGCTGGACAAGATCCGGGCATGCAGAACGCTGAGCACACCCCCAGAAGAAGTG
 GGGTCCCCGACCCAGGCGAACGGTGGCTCCCATAGGACAATCGTACCCCCCGACCTCGT
 AGCAACAGCAATACCGGGGGACCCCTGGCCAGGCCTGGTCCATGAGCAGGGCTCCTCG
 TGCCCTGGCCAGGGTCTCTTCCCTGGCCCCCTCAGTTTCACTTTGGATTTTT
 ATTGTTATTAACATGGACTTGTGT

FIG. 3-34

54712 CAGAGCCCCCCCCGGGGCCGCAGGAGAGGGCCGGGCTGGGCGGATGATGAGGGCCAGT
GAGGCGCCAAGGGAAGGTACCATCAAGTATGACCCCCAAGGAGCTACGGAAGCACCTCAA
CCTAGAGGAGTGGATCTGGAGCAGCTCACCGCCTCTACGACTGCAGGAAGAGGAGAT
CTCAGAACTAGAGATTGACGTGGATGAGCTCTGGACATGGAGAGTGACGATGCTGGC
TTCCAGGGTCAAGGAGCTGCTGGTTACTGTTACAAACCCACAGAGGCCATCTCTGG
[T, C]
CTGCTGGACAAGATCCGGGCCATGCAGAAGCTGAGCACACCCCCAGAAGAAGTGAGGGTCC
CCGACCCAGGCAGCGAACGGTGGCTCCCATAGGACAATCGTACCCCCCGACCTCGTAGAAC
AGCAATACCGGGGGACCCCTGCGGCCAGGCTGGTCCATGAGCAGGGCTCTCGTGC
TGGCCCAAGGGGTCTTCCCTGCCCCCTCAGTTTCCACTTTGGATTTTTATTGTT
ATTAAACTGATGGGACTTTGTGTTTTATATTGACTCTGCGGCACGGGCCCTTAAATAAA
54799 GTATGACCCCCAAGGAGCTACGGAAGCACCTAACCTAGAGGAGTGGATCCTGGAGCAGCT
CACGCGCCTCTACGACTGCCAGGAAGAGGAGATCTCAGAACTAGAGATTGACGTGGATGA
GCTCCTGGACATGGAGAGTGACGATGCTGGCTTCCAGGGTCAAGGAGCTGCTGGTTGA
CTGTTACAAACCCACAGAGGCCATCTCTGGCTGCTGGACAAGATCCGGGCCATGCA
GAAGCTGAGCACACCCCCAGAAGAAGTGAGGGTCCCCGACCCAGGCGAACGGTGGCTCC
[T, C]
AGGACAATCGTACCCCCCGACCTCGTAGAACAGCAATACCGGGGACCCCTGCGGCCAG
GCCTGGTCCATGAGCAGGGCTCTCGTGCCTGGCCAGGGTCTCTTCCCCCTGCC
CTCAGTTTCCACTTTGGATTTTTATTGTTATTAAACTGATGGGACTTTGTGTTTT
ATATTGACTCTGCGGCACGGGCCCTTAAATAAAAGCGAGGTAGGGTACGCCCTTGGTGCAG
CTCAAAAAAAAAAAAAAAATGATTTCCAGGGTCCACATTAGAGTTGAAATTCTGGT
54819 GGAAGCACCTAACCTAGAGGAGTGGATCCTGGAGCAGCTACGCGCCTCTACGACTGCC
AGGAAGAGGAGATCTCAGAACTAGAGATTGACGTGGATGAGCTCTGGACATGGAGAGTG
ACGATGCCCTGGCTCCAGGGTCAAGGAGCTGCTGGTACTGTTACAAACCCACAGAGG
CCTTCATCTCTGGCTGCTGGACAAGATCCGGGCCATGCGAACAGCTGAGCACACCCCCAGA
AGAAGTGAGGGTCCCCGACCCAGGCGAACGGTGGCTCCATAGGACAATCGTACCCCC
[G, A]
ACCTCGTAGCAACAGCAATACCGGGGACCCCTGCGGCCAGGCTGGTCCATGAGCAGGG
CTCCTCGTGCCTGGCCAGGGTCTCTCCCTGCCCTCAGTTTCCACTTTGG
TTTTTTATTGTTATTAAACTGATGGGACTTTGTGTTTTATTGACTCTGCGGCACGG
GCCCTTAAATAAGCGAGGTAGGGTACGCCCTTGGTGCAGCTCAAAAAAAAAAAAAAA
TGATTTCCAGGGTCCACATTAGAGTTGAAATTCTGGTGGGAGAATCTATACCTTGT
55499 TTGTTTCTAATACCTTGTCAATTCAAATATCTTAAATTATTAAAAATATATAT
ACAGTATTGAATGCCACTGTGTCTAGGTACAGTTCAACACCTGGGTTACAGCAGCG
AACAAAATAAAGGTCTTACCCATAGAACATAGATTCTAGCATGGTATCTACTGTATC
ATACAGTAGATAACAATAAGTAAACTATATTGAATATTAGAATGTGGCAGATGCTATGGAA
AAAGAGTCAAGACAAGTAAAGACGATTGTTAGGGTACCCAGTGGTACAGTTGCAATTAAATATGGT
[C, T]
GTCAGAGCAGGCCACTGAGGTGACATGACATTAAAGCATAAACATGGAGGAGGAGGAG
TAAGCCTGAGCTGTCTAGGCTTCCGGGGCAGCCAAGCCATTCCGTGGCACTAGGAGC
TGGTGTTCGATTCCACCTTGATAACTGCATTCTCAAGATATGGAGGGAGTT
TTCTCCTATTGTTTAAGTATTAACTCCAGCTAGTCCAGCCTGTTAGTGTACCTA
ATCTTATAGCAAATATGAGGTACCGGTAACATTATGCCATTCTCACAGAGGCACT

FIG.3-35

56825 ACTGATGGCTAAAGGGTGTGAAAAAGTCAGTGATGCTCCCCTTTACTCCAGATCCT
GTCCCTCCTGGAGCAAGGTTGAGGGAGTAGGTTTGAAAGAGTCCCTTAATATGTGGTGG
ACAGGCCAGGAGTTAGAGAAAGGGCTGGCTTCTGTTACCTGCTACTGGCTCTAGCCAG
CCCAGGGACCACATCAATGTGAGAGGAAGCCTCCACCTCATGTTTCAAACCTTAATACTG
GAGACTGGCTGAGAACTTACGGACAACATCCTTCTGCTGAAACAAACAGTCACAAGCA
[C,A]
AGGAAGAGGCTGGGGACTAGAAAGAGGCCCTGCCCTCTAGAAAGCTCAGATCTGGCTT
CTGTTACTCATACTCGGGTGGGCTCCTTAGTCAGATGCCAAAACATTTCGCTAAAGCT
CGATGGGTTCTGGAGGACAGTGTGGCTTGTACAGGCCCTAGAGTCTGAGGGAGGGAGTG
GGAGTCTCAGCAATCTTGGCTTGGCTTATGGCAACCACTGCTACCCCTCAACATG
CCTGGTTAGGCAGCAGCTGGCTGGGAAGAGGTGGTGGCAGAGTCTCAAAGCTGAGAT

58871 CGTCACCCACCAACCCCTGCCGCACTCCAGCCTTAAACAAGGGCTGTCTAGATATT
CATTTTAACCTACCTCCACCTTGGAAACAATTGCTGAAGGGGAGAGGAATTGCAATGACCA
ACCACCTTGTGGGACGCCCTGCACACCTGTCTTCTGCTTCAACCTGAAAGATTCTGA
TGATGATAATCTGGACACAGAACGCCGGCACGGTGGCTAGCCTGTAATCTCAGCACTT
TGGGAGGCCTCAGCAGGTGGATCACCTGAGATCAAGAGTTGAGAACAGCCTGACCAACA
[T,A]
GGTGAAACCCGTCTACTAAAAATACAAAAATTAGCCAGGTGTGGTGGCACATACCTG
TAATCCCAGCTACTCTGGAGGCTGAGGCAGGAGAACGCTTGAACCCACAAGGCAGAGGT
TGCAGTGAGGCAGATCATGCCATTGCACTCCAGCCTGTGCAACAAGAGCCAAACTCCAT
CTCAAAAAAA

FIG. 3-36

**ISOLATED HUMAN KINASE PROTEINS,
NUCLEIC ACID MOLECULES ENCODING
HUMAN KINASE PROTEINS, AND USES
THEREOF**

FIELD OF THE INVENTION

The present invention is in the field of kinase proteins that are related to the serine/threonine kinase subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel peptides and proteins that effect protein phosphorylation and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

Protein Kinases

Kinases regulate many different cell proliferation, differentiation, and signaling processes by adding phosphate groups to proteins. Uncontrolled signaling has been implicated in a variety of disease conditions including inflammation, cancer, arteriosclerosis, and psoriasis. Reversible protein phosphorylation is the main strategy for controlling activities of eukaryotic cells. It is estimated that more than 1000 of the 10,000 proteins active in a typical mammalian cell are phosphorylated. The high energy phosphate, which drives activation, is generally transferred from adenosine triphosphate molecules (ATP) to a particular protein by protein kinases and removed from that protein by protein phosphatases. Phosphorylation occurs in response to extracellular signals (hormones, neurotransmitters, growth and differentiation factors, etc), cell cycle checkpoints, and environmental or nutritional stresses and is roughly analogous to turning on a molecular switch. When the switch goes on, the appropriate protein kinase activates a metabolic enzyme, regulatory protein, receptor, cytoskeletal protein, ion channel or pump, or transcription factor.

The kinases comprise the largest known protein group, a superfamily of enzymes with widely varied functions and specificities. They are usually named after their substrate, their regulatory molecules, or some aspect of a mutant phenotype. With regard to substrates, the protein kinases may be roughly divided into two groups; those that phosphorylate tyrosine residues (protein tyrosine kinases, PTK) and those that phosphorylate serine or threonine residues (serine/threonine kinases, STK). A few protein kinases have dual specificity and phosphorylate threonine and tyrosine residues. Almost all kinases contain a similar 250-300 amino acid catalytic domain. The N-terminal domain, which contains subdomains I-IV, generally folds into a two-lobed structure, which binds and orients the ATP (or GTP) donor molecule. The larger C terminal lobe, which contains subdomains VI A-XI, binds the protein substrate and carries out the transfer of the gamma phosphate from ATP to the hydroxyl group of a serine, threonine, or tyrosine residue. Subdomain V spans the two lobes.

The kinases may be categorized into families by the different amino acid sequences (generally between 5 and 100 residues) located on either side of, or inserted into loops of, the kinase domain. These added amino acid sequences allow the regulation of each kinase as it recognizes and interacts with its target protein. The primary structure of the kinase domains is conserved and can be further subdivided into 11 subdomains. Each of the 11 subdomains contains specific residues and motifs or patterns of amino acids that

are characteristic of that subdomain and are highly conserved (Hardie, G. and Hanks, S. (1995) *The Protein Kinase Facts Books*, Vol I:7-20 Academic Press, San Diego, Calif.).

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic-ADP-ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The cyclic-AMP dependent protein kinases (PKA) are important members of the STK family. Cyclic-AMP is an intracellular mediator of hormone action in all prokaryotic and animal cells that have been studied. Such hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cyclic-AMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K. J. et al. (1994) *Harrison's Principles of Internal Medicine*, McGraw-Hill, New York, N.Y., pp. 416-431, 1887).

Calcium-calmodulin (CaM) dependent protein kinases are also members of STK family. Calmodulin is a calcium receptor that mediates many calcium regulated processes by binding to target proteins in response to the binding of calcium. The principle target protein in these processes is CaM dependent protein kinases. CaM-kinases are involved in regulation of smooth muscle contraction (MLC kinase), glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM kinase I phosphorylates a variety of substrates including the neurotransmitter related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) *EMBO Journal* 14:3679-86). CaM II kinase also phosphorylates synapsin at different sites, and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. Many of the CaM kinases are activated by phosphorylation in addition to binding to CaM. The kinase may autophosphorylate itself, or be phosphorylated by another kinase as part of a "kinase cascade".

Another ligand-activated protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) *J. Biol. Chem.* 15:8675-81). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

The mitogen-activated protein kinases (MAP) are also members of the STK family. MAP kinases also regulate intracellular signaling pathways. They mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S. E. and Weinberg, R. A. (1993) *Nature* 365:781-783). MAP kinase signaling

pathways are present in mammalian cells as well as in yeast. The extracellular stimuli that activate mammalian pathways include epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, endotoxic lipopolysaccharide (LPS), and pro-inflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1).

PRK (proliferation-related kinase) is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) *J. Biol. Chem.* 271:19402-8). PRK is related to the polo (derived from humans polo gene) family of STKs implicated in cell division. PRK is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation. Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.

The cyclin-dependent protein kinases (CDKs) are another group of STKs that control the progression of cells through the cell cycle. Cyclins are small regulatory proteins that act by binding to and activating CDKs that then trigger various phases of the cell cycle by phosphorylating and activating selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to the binding of cyclin, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue.

Protein tyrosine kinases, PTKs, specifically phosphorylate tyrosine residues on their target proteins and may be divided into transmembrane, receptor PTKs and nontransmembrane, non-receptor PTKs. Transmembrane protein-tyrosine kinases are receptors for most growth factors. Binding of growth factor to the receptor activates the transfer of a phosphate group from ATP to selected tyrosine side chains of the receptor and other specific proteins. Growth factors (GF) associated with receptor PTKs include; epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Non-receptor PTKs lack transmembrane regions and, instead, form complexes with the intracellular regions of cell surface receptors. Such receptors that function through non-receptor PTKs include those for cytokines, hormones (growth hormone and prolactin) and antigen-specific receptors on T and B lymphocytes.

Many of these PTKs were first identified as the products of mutant oncogenes in cancer cells where their activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs, and it is well known that cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Carboneau H and Tonks NK (1992) *Annu. Rev. Cell. Biol.* 8:463-93). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

LIM Domain Kinases

The novel human protein, and encoding gene, provided by the present invention is related to the family of serine/threonine kinases in general, particularly LIM domain kinases (LIMK), and shows the highest degree of similarity to LIMK2, and the LIMK2b isoform (Genbank gi8051618) in particular (see the amino acid sequence alignment of the protein of the present invention against LIMK2b provided in

FIG. 2). LIMK proteins generally have serine/threonine kinase activity. The protein of the present invention may be a novel alternative splice form of the art-known protein provided in Genbank gi805161 ; however, the structure of the gene provided by the present invention is different from the art-known gene of gi8051618 and the first exon of the gene of the present invention is novel, suggesting a novel gene rather than an alternative splice form. Furthermore, the protein of the present invention lacks an LIM domain relative to gi8051618. The protein of the present invention does contain the kinase catalytic domain.

Approximately 40 LIM proteins, named for the LIM domains they contain, are known to exist in eukaryotes. LIM domains are conserved, cystein-rich structures that contain 2 zinc fingers that are thought to modulate protein-protein interactions. LIMK1 and LIMK2 are members of a LIM subfamily characterized by 2 N-terminal LIM domains and a C-terminal protein kinase domain. LIMK1 and LIMK2 mRNA expression varies greatly between different tissues. The protein kinase domains of LIMK1 and LIMK2 contain a unique sequence motif comprising Asp-Leu-Asn-Ser-His-Asn in subdomain VIB and a strongly basic insert between subdomains VII and VIII (Okano et al., *J. Biol. Chem.* 270 (52), 31321-31330 (1995)). The protein kinase domain present in LIMKs is significantly different than other kinase domains, sharing about 32% identity.

LIMK is activated by ROCK (a downstream effector of Rho) via phosphorylation. LIMK then phosphorylates cofilin, which inhibits its actin-depolymerizing activity, thereby leading to Rho-induced reorganization of the actin cytoskeleton (Maekawa et al., *Science* 285: 895-898, 1999).

The LIMK2a and LIMK2b alternative transcript forms are differentially expressed in a tissue-specific manner and are generated by variation in transcriptional initiation utilizing alternative promoters. LIMK2a contains 2 LIM domains, a PDZ domain (a domain that functions in protein-protein interactions targeting the protein to the submembranous compartment), and a kinase domain; whereas LIMK2b just has 1.5 LIM domains. Alteration of LIMK2a and LIMK2b regulation has been observed in some cancer cell lines (Osada et al., *Biochem. Biophys. Res. Commun.* 229: 582-589, 1996).

For a further review of LIMK proteins, see Nomoto et al., *Gene* 236 (2), 259-271 (1999).

Kinase proteins, particularly members of the serine/threonine kinase subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown members of this subfamily of kinase proteins. The present invention advances the state of the art by providing previously unidentified human kinase proteins that have homology to members of the serine/threonine kinase subfamily.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human kinase peptides and proteins that are related to the serine/threonine kinase subfamily, as well as allelic variants and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate kinase activity in cells and tissues that express the kinase. Experimental data as provided in FIG. 1

indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland.

DESCRIPTION OF THE FIGURE SHEETS

FIG. 1 provides the nucleotide sequence of a cDNA molecule that encodes the kinase protein of the present invention. (SEQ ID NO:1) In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland.

FIG. 2 provides the predicted amino acid sequence of the kinase of the present invention. (SEQ ID NO:2) In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIG. 3 provides genomic sequences that span the gene encoding the kinase protein of the present invention. (SEQ ID NO:3) In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. As illustrated in FIG. 3, SNPs were identified at 42 different nucleotide positions.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a kinase protein or part of a kinase protein and are related to the serine/threonine kinase subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human kinase peptides and proteins that are related to the serine/threonine kinase subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these kinase peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the kinase of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known kinase proteins of the serine/threonine kinase subfamily and the expression pattern observed. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. The art has clearly established the commercial importance of

members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known serine/threonine kinase family or subfamily of kinase proteins.

10 Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as 15 being members of the kinase family of proteins and are related to the serine/threonine kinase subfamily (protein sequences are provided in FIG. 2, transcript/cDNA sequences are provided in FIG. 1 and genomic sequences are provided in FIG. 3). The peptide sequences provided in FIG. 20 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in FIG. 3, will be referred herein as the kinase peptides of the present invention, kinase peptides, or peptides/proteins of the present invention.

25 The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprise the amino acid sequences of the kinase peptides disclosed in the FIG. 2, (encoded by the nucleic acid 30 32 genomic sequence shown in FIG. 1, transcript/cDNA or FIG. 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or 35 "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" 45 includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the kinase peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated kinase peptide can be purified from cells that 65 naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as

provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. For example, a nucleic acid molecule encoding the kinase peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in FIG. 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in FIG. 1 (SEQ ID NO:1) and the genomic sequences provided in FIG. 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in FIG. 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in FIG. 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in FIG. 1 (SEQ ID NO:1) and the genomic sequences provided in FIG. 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in FIG. 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in FIG. 1 (SEQ ID NO:1) and the genomic sequences provided in FIG. 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the kinase peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The kinase peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a kinase peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the kinase peptide. "Operatively linked" indicates that the kinase peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the kinase peptide.

In some uses, the fusion protein does not affect the activity of the kinase peptide per se. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant kinase peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel et al., *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A kinase peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the kinase peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the kinase peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A. M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D. W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A. M., and Griffin, H. G.,

eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., et al., *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a NWS-gapDNA.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the kinase peptides of the present invention as well as being encoded by the same genetic locus as the kinase peptide provided herein. The gene encoding the novel kinase protein of the present invention is located on a genome component that has been mapped to human chromosome 22 (as indicated in FIG. 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

Allelic variants of a kinase peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the kinase peptide as well as being encoded by the same genetic locus as the kinase peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in FIG. 3, such as the genomic sequence mapped to the reference human. The gene encoding the novel kinase protein of the present invention is located on a genome component that has been mapped to human chromosome 22 (as indicated in FIG. 3), which is supported by multiple lines of evidence, such as STS and BAC map data. As used herein, two proteins (or a region of

the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a kinase peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

FIG. 3 provides information on SNPs that have been found in the gene encoding the kinase protein of the present invention. SNPs were identified at 42 different nucleotide positions. Some of these SNPs, which are located outside the ORF and in introns, may affect gene transcription.

Paralogs of a kinase peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the kinase peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a kinase peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a kinase peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the kinase peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a kinase peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the kinase peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the kinase peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a kinase peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie et al., *Science* 247:1306-1310 (1990).

Variant kinase peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind substrate, ability to phosphorylate substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. FIG. 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham et al., *Science* 244:1081-1085 (1989)), particularly using the results provided in FIG. 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as kinase activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith et al., *J. Mol. Biol.* 224:899-904 (1992); de Vos et al. *Science* 255:306-312 (1992)).

The present invention further provides fragments of the kinase peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in FIG. 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a kinase peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the kinase peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the kinase peptide, e.g., active site, a transmembrane domain or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in FIG. 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in kinase peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in FIG. 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristylation, oxidation, pro-

teolytic processing, phosphorylation, prenylation, racemization, selenylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins—Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifert et al. (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan et al. (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

Accordingly, the kinase peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature kinase peptide is fused with another compound, such as a compound to increase the half-life of the kinase peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature kinase peptide, such as a leader or secretory sequence or a sequence for purification of the mature kinase peptide or a pro-protein sequence.

30 Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as, for example, in a kinase-effector protein interaction or kinase-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, kinases isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the kinase. Experimental data as provided in FIG. 1 indicates that the kinase proteins of the present invention are expressed in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant

brain, and thyroid gland, as indicated by virtual northern blot analysis. In addition, PCR-based tissue screening panels indicate expression in fetal brain. A large percentage of pharmaceutical agents are being developed that modulate the activity of kinase proteins, particularly members of the serine/threonine kinase subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in FIG. 1. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. Such uses can readily be determined using the information provided herein, that which is known in the art, and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to kinases that are related to members of the serine/threonine kinase subfamily. Such assays involve any of the known kinase functions or activities or properties useful for diagnosis and treatment of kinase-related conditions that are specific for the subfamily of kinases that the one of the present invention belongs to, particularly in cells and tissues that express the kinase. Experimental data as provided in FIG. 1 indicates that the kinase proteins of the present invention are expressed in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant brain, and thyroid gland, as indicated by virtual northern blot analysis. In addition, PCR-based tissue screening panels indicate expression in fetal brain.

The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the kinase, as a biopsy or expanded in cell culture. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the kinase protein.

The polypeptides can be used to identify compounds that modulate kinase activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the kinase. Both the kinases of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the kinase. These compounds can be further screened against a functional kinase to determine the effect of the compound on the kinase activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the kinase to a desired degree.

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the kinase protein and a molecule that normally interacts with the kinase protein, e.g. a substrate or a component of the signal pathway that the kinase protein normally interacts (for example, another kinase). Such assays typically include the steps of combining the kinase protein with a candidate compound under conditions that allow the kinase protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the kinase protein and the target, such as any of the associated effects of signal

transduction such as protein phosphorylation, cAMP turnover, and adenylate cyclase activation, etc.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam et al., *Nature* 354:82-84 (1991); Houghten et al., *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L-configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang et al., *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for substrate binding. Other candidate compounds include mutant kinases or appropriate fragments containing mutations that affect kinase function and thus compete for substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) kinase activity. The assays typically involve an assay of events in the signal transduction pathway that indicate kinase activity. Thus, the phosphorylation of a substrate, activation of a protein, a change in the expression of genes that are up- or down-regulated in response to the kinase protein dependent signal cascade can be assayed.

Any of the biological or biochemical functions mediated by the kinase can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly FIG. 2. Specifically, a biological function of a cell or tissues that expresses the kinase can be assayed. Experimental data as provided in FIG. 1 indicates that the kinase proteins of the present invention are expressed in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant brain, and thyroid gland, as indicated by virtual northern blot analysis. In addition, PCR-based tissue screening panels indicate expression in fetal brain.

Binding and/or activating compounds can also be screened by using chimeric kinase proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a substrate-binding region can be used that interacts with a different substrate than that which is recognized by the native kinase. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the kinase is derived.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the kinase (e.g. binding part-

ners and/or ligands). Thus, a compound is exposed to a kinase polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble kinase polypeptide is also added to the mixture. If the test compound interacts with the soluble kinase polypeptide, it decreases the amount of complex formed or activity from the kinase target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the kinase. Thus, the soluble polypeptide that competes with the target kinase region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the kinase protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S -labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of kinase-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a kinase-binding protein and a candidate compound are incubated in the kinase protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the kinase protein target molecule, or which are reactive with kinase protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the kinases of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of kinase protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the kinase pathway, by treating cells or tissues that express the kinase. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. These methods of treatment include the steps of administering a modulator of

kinase activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the kinase proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Pat. No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693/1696; and Brent WO94110300), to identify other proteins, which bind to or interact with the kinase and are involved in kinase activity. Such kinase-binding proteins are also likely to be involved in the propagation of signals by the kinase proteins or kinase targets as, for example, downstream elements of a kinase-mediated signaling pathway. Alternatively, such kinase-binding proteins are likely to be kinase inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a kinase protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a kinase-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the kinase protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a kinase-modulating agent, an antisense kinase nucleic acid molecule, a kinase-specific antibody, or a kinase-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The kinase proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. The method involves contacting a biological sample with a compound capable of interacting with the kinase protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A bio-

logical sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered kinase activity in cell-based or cell-free assay, alteration in substrate or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected in vivo in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M. W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the kinase protein in which one or more of the kinase functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other substrate-binding regions that are

more or less active in substrate binding, and kinase activation. Accordingly, substrate dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. Accordingly, methods for treatment include the use of the kinase protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')₂, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, *Antibodies*, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in FIG. 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the kinase proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or kinase/binding partner interaction. FIG. 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see FIG. 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody

to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliflavin, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luciferin; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in FIG. 1 indicates that the kinase proteins of the present invention are expressed in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant brain, and thyroid gland, as indicated by virtual northern blot analysis. In addition, PCR-based tissue screening panels indicate expression in fetal brain. Further, such antibodies can be used to detect protein in situ, in vitro, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic

proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the kinase peptide to a binding partner such as a substrate. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See FIG. 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a kinase peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the kinase peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by

recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in FIG. 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in FIG. 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in FIG. 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in FIG. 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in FIG. 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in FIG. 2, SEQ ID NO:2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In FIGS. 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (FIG. 3) and cDNA/transcript sequences (FIG. 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in FIGS. 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of

a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the kinase peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the kinase proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in FIGS. 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in FIG. 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. The gene encoding the novel kinase protein of the present invention is located on a genome component that has been mapped to human chromosome 22 (as indicated in FIG. 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

FIG. 3 provides information on SNPs that have been found in the gene encoding the kinase protein of the present invention. SNPs were identified at 42 different nucleotide positions. Some of these SNPs, which are located outside the ORF and in introns, may affect gene transcription.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2×SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in FIG. 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in FIG. 2. As illustrated in FIG. 3, SNPs were identified at 42 different nucleotide positions.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter in situ expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of in situ hybridization methods. The gene encoding the novel kinase protein of the present invention is located on a genome component that has been mapped to human chromosome 22 (as indicated in FIG. 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in FIG. 1 indicates that the kinase proteins of the present invention are expressed in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant brain, and thyroid gland, as indicated by virtual northern blot analysis. In addition, PCR-based tissue screening panels indicate expression in fetal brain. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in kinase protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detecting DNA includes Southern hybridizations and in situ hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a kinase protein, such as by measuring a level of a kinase-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a kinase gene has been mutated. Experimental data as provided in FIG. 1 indicates that the kinase proteins of the present invention are expressed in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant brain, and thyroid gland, as indicated by

virtual northern blot analysis. In addition, PCR-based tissue screening panels indicate expression in fetal brain.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate kinase nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the kinase gene, particularly biological and pathological processes that are mediated by the kinase in cells and tissues that express it. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland. The method typically includes assaying the ability of the compound to modulate the expression of the kinase nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired kinase nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the kinase nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for kinase nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the kinase protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of kinase gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of kinase mRNA in the presence of the candidate compound is compared to the level of expression of kinase mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate kinase nucleic acid expression in cells and tissues that express the kinase. Experimental data as provided in FIG. 1 indicates that the kinase proteins of the present invention are expressed in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant brain, and thyroid gland, as indicated by virtual northern blot analysis. In addition, PCR-based tissue screening panels indicate expression in fetal brain. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for kinase nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the kinase nucleic acid expression in the cells and tissues that express the protein. Experimental data as provided in FIG. 1 indicates expression in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant and fetal brain, and thyroid gland.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the kinase gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in kinase nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in kinase genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the kinase gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the kinase gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a kinase protein.

Individuals carrying mutations in the kinase gene can be detected at the nucleic acid level by a variety of techniques. FIG. 3 provides information on SNPs that have been found in the gene encoding the kinase protein of the present invention. SNPs were identified at 42 different nucleotide positions. Some of these SNPs, which are located outside the ORF and in introns, may affect gene transcription. The gene encoding the novel kinase protein of the present invention is located on a genome component that has been mapped to human chromosome 22 (as indicated in FIG. 3), which is supported by multiple lines of evidence, such as STS and BAC map data. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al., *Science* 241:1077-1080 (1988); and Nakazawa et al., *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya et al., *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal

genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a kinase gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Pat. No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant kinase gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C. W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen et al., *Adv. Chromatogr.* 36:127-162 (1996); and Griffin et al., *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers et al., *Science* 230:1242 (1985)); Cotton et al., *PNAS* 85:4397 (1988); Saleeba et al., *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita et al., *PNAS* 86:2766 (1989); Cotton et al., *Mutat. Res.* 285:125-144 (1993); and Hayashi et al., *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers et al., *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the kinase gene in an individual in order to select an appropriate compound or dosage regimen for treatment. FIG. 3 provides information on SNPs that have been found in the gene encoding the kinase protein of the present invention. SNPs were identified at 42 different nucleotide positions. Some of these SNPs, which are located outside the ORF and in introns, may affect gene transcription.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control kinase gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene

involved in transcription, preventing transcription and hence production of kinase protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into kinase protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of kinase nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired kinase nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the kinase protein, such as substrate binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in kinase gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired kinase protein to treat the individual.

The invention also encompasses kits for detecting the presence of a kinase nucleic acid in a biological sample. Experimental data as provided in FIG. 1 indicates that the kinase proteins of the present invention are expressed in humans in teratocarcinoma, ovary, testis, nervous tissue, bladder, infant brain, and thyroid gland, as indicated by virtual northern blot analysis. In addition, PCR-based tissue screening panels indicate expression in fetal brain. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting kinase nucleic acid in a biological sample; means for determining the amount of kinase nucleic acid in the sample; and means for comparing the amount of kinase nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect kinase protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in FIGS. 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in U.S. Pat. No. 5,837,832, Chee et al., PCT application WO95/11995 (Chee et al.), Lockhart, D. J. et al. (1996; *Nat. Biotech.* 14: 1675-1680) and Schena, M. et al. (1996; *Proc. Natl. Acad. Sci.* 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et al., U.S. Pat. No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia

viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, Streptomyces, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as Drosophila, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith et al., *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, Mass.) and pRITS (Pharmacia, Piscataway, N.J.) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amano et al., *Gene* 69:301-315 (1988)) and pET 11 d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada et al., *Nucleic Acids Res.* 20:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYEpSec1 (Baldari, et al., *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan et al., *Cell* 30:933-943(1982)), pJRY88 (Schultz et al., *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, Calif.).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow et al., *Virology* 170:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman et al., *EMBO J.* 6:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the

recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as kinases, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with kinases, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of Vectors and Host Cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a kinase protein or peptide that can be further purified to produce desired amounts of kinase protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the kinase protein or kinase protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native kinase protein is useful for assaying compounds that stimulate or inhibit kinase protein function.

Host cells are also useful for identifying kinase protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant kinase protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native kinase protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for

studying the function of a kinase protein and identifying and evaluating modulators of kinase protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the kinase protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence (s) can be operably linked to the transgene to direct expression of the kinase protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Pat. No. 4,873,191 by Wagner et al. and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, e.g., Lasko et al. *PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman et al. *Science* 251:1351-1355 (1991). If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recom-

binase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. et al. *Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect substrate binding, kinase protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* kinase protein function, including substrate interaction, the effect of specific mutant kinase proteins on kinase protein function and substrate interaction, and the effect of chimeric kinase proteins. It is also possible to assess the effect of null mutations, that is, mutations that substantially or completely eliminate one or more kinase protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 4

<210> SEQ ID NO 1
<211> LENGTH: 2320
<212> TYPE: DNA
<213> ORGANISM: Human

<400> SEQUENCE: 1

ccccagggcgc cgtaggcggt gcatcccggt cgcgcctggg gctgtggct tcccgccgcct	60
gaggcgccgg cggcaggagc tgaggggagt tgtagggAAC tgaggggAGC tgctgtgtcc	120
cccgccctct cctccccatt tccgcgcgtcc cgggaccatg tccgcgcgtgg cgggtgaaga	180
tgtctggagg tgtccaggct gtggggacca cattgtccca agccagatAT ggtacaggac	240

-continued

tgtcaacgaa	acctggcacg	gctttgttcc	ccgggtgaaag	tgtatgcgcag	cctggaccac	300
cccaatgtgc	tcaagttcat	tggtgtgtcg	tacaaggata	agaagctgaa	cctgtctgaca	360
gagtagatcc	aggggggcac	actgaaggac	tttctgcgc	gtatggatcc	gttccccctgg	420
cagcagaagg	tcaggtttgc	caaaggaaatc	gcctccggaa	tggacaagac	tgtgggttg	480
gcagacttg	ggctgtcacg	gctcatatgc	gaagagagga	aaagggcccc	catggagaag	540
gccaccacca	agaaacgcac	cttgcgc	aaacgaccgc	agaagcgct	cacgggttg	600
ggaaacccct	actggatggc	ccctgagatg	ctgaacggaa	agagctatga	tgagacggtg	660
gatatcttct	cctttggat	cggtctctgt	gagatcattg	ggcagggtga	tgcagatcct	720
gactgccttc	cccgaaacact	ggactttggc	ctcaacgtga	agctttctg	ggagaagtt	780
gttcccacag	attgtccccc	ggccttc	ccgcgtggcc	ccatctgt	cagactggag	840
cctgagagca	gaccagcatt	ctcgaaattg	gaggactct	ttgaggccct	ctccctgtac	900
ctggggggagc	tgggcatccc	gctgcctgc	gagctggagg	agttggacca	cactgtgagc	960
atgcagtacg	gcctgaccgg	ggactcacct	ccctagccct	ggcccgccc	cctgcagggg	1020
ggtgttctac	agccagcatt	gcccctctgt	gccccattcc	tgctgtgagc	agggccgtcc	1080
gggcttc	tggattggcg	gaatgtttag	aagcagaaaca	aaccattct	attacctccc	1140
caggaggcaa	gtgggcgcag	caccaggaa	atgtatctcc	acagttctg	gggccttagtt	1200
actgtctgt	aatccaaatac	ttgcctgaaa	gctgtgaaga	agaaaaaaac	ccctggcctt	1260
tggggcagga	ggaatcttt	actgaatcc	accaggaaac	tccctggcg	tggattgtgg	1320
gaggcttctg	cttacactaa	tcagcgtac	ctggacatgc	tggcaggat	cccagggtga	1380
acctgcctgt	gaactctgaa	gtcacttagtc	cagctgggtg	caggaggact	tcaagtgtgt	1440
ggacgaaaga	aagactgtat	gctcaaagggg	tgtaaaaaaag	tcaatgtatgc	tcccccttc	1500
tactccatag	cctgtcttc	ctggagcaag	gttggggag	taggtttga	agagtccctt	1560
aatatgttgt	ggaacaggcc	aggagttaga	gaaagggtg	gcttctgtt	acctgctcac	1620
tggctctagc	cagcccagg	accacatcaa	tgtgagagga	agcccccacc	tcatgttttc	1680
aaacttaata	ctggagactg	gctgagaact	tacggacaac	atcctttctg	tctgaaacaa	1740
acagtca	acaa	gcacaggaag	aggctgggg	actagaaaga	ggccctgccc	1800
tcagatcttgc	gtttctgtt	ctcatactcg	ggggcgtcc	ttatgtcagat	gcctaaaaca	1860
ttttgcctaa	agctcgatgg	gttctggagg	acagtgtggc	ttgtcacagg	cctagatgt	1920
gagggggggg	agtggggatgc	tcaatgtat	cttggcttgc	gcttcatggc	aaccactgt	1980
cacccttcaa	catgcgttgt	ttaggcagca	gcttgggtcg	ggaagagggt	gtggcagagt	2040
ctcaaaatgt	agatgtctgt	agagatagct	ccctgagctg	ggccatctga	cttctaccc	2100
ccatgtttgc	tctccaaat	cattagctcc	tggcagcat	cctccgtgac	cacatgtca	2160
ggtaactggaa	aacactccatc	ttggctccca	gagctttagg	aactcttcat	cacaactaga	2220
tttgccttct	ctaaatgttct	atgagcttgc	accatattta	ataaaattggg	aatgggttg	2280
gggttataaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa			2320

<210> SEQ ID NO 2

<211> LENGTH: 255

<212> TYPE: PRT

<213> ORGANISM: Human

<400> SEQUENCE: 2

Met Val Gln Asp Cys Gln Arg Asn Leu Ala Arg Leu Leu Leu Pro Val

-continued

1	5	10	15
Lys Val Met Arg Ser Leu Asp His Pro Asn Val Leu Lys Phe Ile Gly			
20	25	30	
Val Leu Tyr Lys Asp Lys Lys Leu Asn Leu Leu Thr Glu Tyr Ile Glu			
35	40	45	
Gly Gly Thr Leu Lys Asp Phe Leu Arg Ser Met Asp Pro Phe Pro Trp			
50	55	60	
Gln Gln Lys Val Arg Phe Ala Lys Gly Ile Ala Ser Gly Met Asp Lys			
65	70	75	80
Thr Val Val Val Ala Asp Phe Gly Leu Ser Arg Leu Ile Val Glu Glu			
85	90	95	
Arg Lys Arg Ala Pro Met Glu Lys Ala Thr Thr Lys Lys Arg Thr Leu			
100	105	110	
Arg Lys Asn Asp Arg Lys Lys Arg Tyr Thr Val Val Gly Asn Pro Tyr			
115	120	125	
Trp Met Ala Pro Glu Met Leu Asn Gly Lys Ser Tyr Asp Glu Thr Val			
130	135	140	
Asp Ile Phe Ser Phe Gly Ile Val Leu Cys Glu Ile Ile Gly Gln Val			
145	150	155	160
Tyr Ala Asp Pro Asp Cys Leu Pro Arg Thr Leu Asp Phe Gly Leu Asn			
165	170	175	
Val Lys Leu Phe Trp Glu Lys Phe Val Pro Thr Asp Cys Pro Pro Ala			
180	185	190	
Phe Phe Pro Leu Ala Ala Ile Cys Cys Arg Leu Glu Pro Glu Ser Arg			
195	200	205	
Pro Ala Phe Ser Lys Leu Glu Asp Ser Phe Glu Ala Leu Ser Leu Tyr			
210	215	220	
Leu Gly Glu Leu Gly Ile Pro Leu Pro Ala Glu Leu Glu Leu Asp			
225	230	235	240
His Thr Val Ser Met Gln Tyr Gly Leu Thr Arg Asp Ser Pro Pro			
245	250	255	

<210> SEQ ID NO 3

<211> LENGTH: 59065

<212> TYPE: DNA

<213> ORGANISM: Human

<400> SEQUENCE: 3

tcatccttgc gcaggggcca tgctaacctt ctgtgtctca gtccaaattt aatgtatgtg	60
ctgctgaagc gagagtagcca gaggttttt tgatggcagt gacttgaact tatttaaaag	120
ataaggagga gccagtgagg gagaggggtg ctgtaaagat aactaaaatgc gacttcttc	180
taagaagtaa gatggaatgg gatccagaac aggggtgtca taccgagtag cccagccctt	240
gttccgtgga cactggggag tctaaccctt agctgagata gcttgcagtggatgagcc	300
agctgagtagc agcagatagg gaaaagaagc caaaaatctg aagttagggct ggggtgaaagg	360
acagggaaagg gctagagaga catttggaaa gtgaaaccag gtggatatga gaggagagag	420
tagagggct tgatccggg tctttcatgc ttaacccaaa gcaggtacta aagtatgtgt	480
tgattgaatg tctttgggtt tctcaagact ggagaaagca gggcaagctc tggagggtat	540
ggcaataaca agttatcttg aatatccctca tggtggaaaag tcctgtatct gtttgaattt	600
tggaaaataga aatcattcag agccaagaga ttgaattgtt gagtaagtgg gtggcaggt	660
tacagactta atttgggtt aaaaatgtaaa aacaagaac aagggtgtggc tctaaaataa	720

-continued

-continued

gaaaacaagct	tctgagttga	gaatttcaat	cttaggggtgg	ggaaaggaaat	gtaccaaaggaa	3180
agagctcatg	acccaaacctc	aagtgtggcc	ccctgttgcacc	caggtaaat	tggaagagcc	3240
ataaaatggc	cagctggagg	cagggtgggg	ggatggagagg	agcccttcc	agggttgtcc	3300
cataccctc	actttatggg	tgagggaaact	gaggcccagg	aagagtact	ttctgtggc	3360
tgcaactacag	attatgcagg	tacttcaaga	gttggggta	ttcttattttt	attttatttt	3420
attttattttt	attttattttt	atttttagag	agggatttctt	gctgttgc	cccagggttgt	3480
gcagtgggtgc	aatctcggt	cactgcaatc	tctgcctgt	gggttcaagt	gattttctg	3540
ccttagcttc	ctgagtagct	gagatgacag	gcacccgtca	ccatgcgcag	ctaattttt	3600
tattttagt	gagacggggg	tttcaacatg	ttgggtcaggc	ttgggttggaa	ctctgttgc	3660
caaatgtgc	acccacccctcg	accccccataa	gtgctggaaat	tacaggcg	aaccactgtg	3720
cccagccaaag	agttgtttttt	agtgtgggttgc	gcagagccag	ctcttccttc	accacaggat	3780
gcctccctag	gttccctactt	tttgggtacta	gttgggttata	tagctatatt	attattatttt	3840
ttattattat	tattattattt	attattgaga	cagagtctcg	ctctgttgc	cagggttgt	3900
tacagtgggt	cgatcccggtt	ctcactgca	cctctgttgc	ccgagttca	gcagtttcc	3960
tgccctcagcc	cccccgagtag	gtgggactac	aggcgccctgc	caccacaccc	ggctaaatttt	4020
tgtatttta	gttagagacgg	gttttccactt	tggttgc	gtgggttgg	agctcttgc	4080
ctcaggttaag	tgctagaatc	acaggcggt	accactgcgc	ccagccaaaga	gttggggta	4140
gtgtgggtgg	cagagccagc	tcttcctcac	cacagggtgc	ctcccttaggt	ttctactttt	4200
tgttactagc	tttattatag	ctacattatt	attattattt	tttattattat	tgagacagag	4260
tctcgctctg	tcgcccaggc	ttgggttacag	tgatgttgc	ttgggttact	gcaacccctg	4320
ccccccgagt	tcaagcaatt	ctccctgttgc	agcccccata	gttaggtggaa	ctccaggcac	4380
ctgcccaccc	gcccagctaa	tttttgtatt	tttagtagag	gcggggtttc	accttgcgttgg	4440
ccaggctgtt	ctcaaaactcc	tgacccctagg	tgatccgcct	gcctcgcc	cccaaaatgt	4500
tgggattaca	ggcatgagcc	accgcgcctc	gcctataatgc	acattttttt	tgtaggcagc	4560
tcaatgttctt	aaaaattata	cagacttca	atcagatttgc	ttctgttgc	ctgaggctca	4620
gtttcttcat	ctggaaaatg	gtgttata	atcttggta	gattgttgc	aaaaatata	4680
gcagtgttac	cagtacatgg	tagacacccca	gtgtatgggtt	attcttgc	cccatcggt	4740
tggaaatttcc	aaagggtggaa	acttgcctt	atattttca	caacgtaaaa	tagttgaaat	4800
tttgggtgg	aaagaagagc	gttccactcc	agagggttgg	tgggttgc	tggccccc	4860
ggctgttgc	gttagggctg	tgcctatata	ctgagaatga	gtatgttgc	gcaggccact	4920
tgtgtgttgc	atccctatgt	ctgcacatag	ctcttgcgtt	aaaatcc	tgtgttgcata	4980
ccaaatgtt	gatgttgcct	ttaaacactt	gccttcc	tggaaaccat	atagggttgc	5040
ggcctggaga	cgtctggcc	ctggaaaggt	tggaaagcg	ccatcattat	tatcccttcc	5100
tttcagctat	aaactcagac	tctcaatgt	tttgcgttgc	tcttatttgc	ttgggttgc	5160
ccccttttac	tcccaggaa	gttggattctg	tctttctgt	tccatattgt	atgacaggag	5220
cagagaatgt	cagactgtt	agggacccca	tagttaaac	ctttgggtgg	tcctttcatt	5280
ttatagctgg	gactaataag	taacgttcaaa	acccaaatgc	ttcacagatt	gggttgc	5340
ttggcatgtt	acccatatgt	tcatattttt	gctgtttcc	tatgtgtat	aatattttt	5400
atccaaaata	agcaggacag	gttagagccaa	gttaatcttt	ggaatttctg	gattcttca	5460

-continued

gagctaaaaa	acttcagaac	tagaagaaac	cacccactat	atggtataac	ccattcatat	5520
cacagatgag	gcctgaaacc	aaaaagactt	gctcaggcca	tggatgacaa	gagctggccc	5580
tagcaactgaa	ctcttgggtc	atttttaggt	ctagtcagat	gctagcttgt	tagctctgtg.	5640
cgtgcgtgtg	tgtgtgtgt	tgtgtgtgt	tgtgtgagat	agagacagaa	agataacata	5700
tgtacacaaa	tacataaaga	ggaagttagac	acgtagcat	ggtagataag	agtacaggca	5760
ggccaggcgt	ggtggctcac	gcctgtatc	ccagcacttt	gggaggccaa	ggcagggtgga	5820
tcacacctgagg	tcaggaattc	gagaccagcc	tgaccaacat	ggtgaaaccc	catctctact	5880
aaatacagaa	aaaaatttgc	ttggcatgtt	ggcacatgcc	tgtatccca	gctacttggg	5940
aagctgaagc	aggagaatcg	cttgaatccg	ggaagcagaa	gttgcagtga	gccgagattt	6000
tgccattaca	gtctagccctg	ggcacaacaaga	gggaaactcc	atcgaaaaaa	aacaaccacc	6060
accaagagta	caggctatgg	aatgagacta	tggttttaaa	tcctggcttt	gcaatttatt	6120
aactagccct	aagtgcactc	cctgagcttc	aggcaccaat	ctgtaaaaatg	aggataagaa	6180
tattactcat	gccacatggt	tgttagggag	gattaaatgt	gataacctat	ataaaagtggc	6240
tagcatagca	tctgacatata	agaaaaactct	taataggcc	ggacgtggtg	gcttatgcct	6300
gtaaatcctag	cactctggga	ggccggaggca	gaaggatcgc	ttgagccat	gagcccgagga	6360
gtttgagacc	agcctggccca	acatggcaaa	actccaccc	tacaaaaaat	acaaaaatata	6420
tagccaggcg	tgtatggcaca	cacctgttgt	cccagctact	tggaaagctg	aggagcgtat	6480
attacctgag	cccagggata	tcaaggctgt	agtgagctgt	gatcatgcc	ctgtactcca	6540
tccagctggg	ggacagagtg	aaacccctgt	ctcaaaacaa	aacaaatgaa	aaaaaaaaacc	6600
cctaataatc	agtaactgtc	actttatatt	atgttgcag	tgtgtgtcta	tatacaccta	6660
tatgtataca	tttctcttat	tacacattca	ttggtagatct	gatgtggagc	cccaggatt	6720
aagggcaact	ttgaactacc	ctgacacaat	caagccaaat	atcattcccg	tggaggaagt	6780
agagtatcta	ggttctgtct	cctagttgca	gctttaccc	gaggacagag	actctaattcc	6840
agctgtgtc	aaggaggcaca	tctcctgact	tctgagcttt	cccctggtaa	attcaaactg	6900
gatgtcacgg	cgccttcaga	tagagcctgg	taatttgccc	tggggagagt	gactgtcttt	6960
tggatctaata	ttgacttttt	ccccagttgg	aggaaaatct	tcagggctag	gaaggattgt	7020
atttgcgtca	ccccagagat	aacctgggtt	ttgaggaaca	tggggcatca	acctgaatgg	7080
tcttgcataa	tctctccac	cccaagcttc	cagtgtttct	ctgtatgaaatt	tagatgtac	7140
gagtagtgc	ggcctgtgg	gaggaggact	ctccctctgt	gctactcaga	gaaattcatt	7200
cttcaaggcc	cccttccagc	cttgccttta	cccaagctgg	ctacagttac	aataaaggaa	7260
atgacttttc	ttctccccc	ccccagtcac	ctttgttttc	ctagtcacag	ggtggggctg	7320
gatattgaat	ggagaaattt	ctggggtcca	tcctaaactc	ctccccctat	cttccttta	7380
cattacccca	ttcttctgtc	tgcagccaca	tcctataatc	tgcctctgtt	agccctccga	7440
cagaccctca	ggtgcccagg	acaacaggaa	gctacttaaa	gctggAACCT	cagactgtgc	7500
atggaggcc	agtgacaaaa	ctgaaaggtag	ctctgtcagt	aatgtgtctg	gtgcgatgt	7560
gcagctggcc	agaatctttt	ggatctctgt	gacatatggc	tgtactgtcc	tcccaagcc	7620
tcccaacagg	cctctttttt	ttcttttttt	ttttttttt	ttttttttt	ttttttttt	7680
tctttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7740
aaatcggtaa	ctggtagtga	tcaattactt	gtaaacacta	ttgtacttgg	accagcccag	7800
taggcctttt	ttaaaaactct	gagttaccc	tctttccctt	ccttgcacgag	tgccatataat	7860

-continued

tctgtatctg gggcaatcct ttctgtatgtt ctctggaccc ggctctctct ccttaggaga 7920
 ggcaggaga gtagccagag agcatgtcat ttgttagctga ggtaaagtg tggagctatc 7980
 aatggtgacc tggccttcttgc gcatgttagc aagccagagg accttgcacaa ctttttgtat 8040
 gattgtccgt tcacccctgtat caaagggtgtt tggcttagga ggagggaaaga aaagctaccc 8100
 ctatttagtct ttagtggcccc aegctgggtc tctattgttgc ttagcagcat 8160
 tatacagaagg aaatccacc gctcttaagg ctccctggaa ctttcaggac ttcccttctc 8220
 aggattgcaa acataagact attttagtgc ttactttgtat aagcggtaa ctaataccta 8280
 tactctggaa aagggtcaat gcagatgaa gactgtggtc actgcatacg gcaacagacc 8340
 atttccgcta attttagtgc ctccagaa gccagtgaa gaaataacaca cgtagcaacc 8400
 agagactgtg ttgtatatgt ttggctgaca gcagggtact ttctgtatgttgc 8460
 cattcatttt ctctcccttc atccccatct aagcaagcct ggtagaatca taattacagt 8520
 aataggtacc acttatttgc tactctgtgc cagacaccct cctgagcata cgacatgc 8580
 agcacatcta atcccttacaa tgacttaataa aaatgttagtgc ttagtcttac ctacttcgag 8640
 aatagggaaa tggaggttac ttgtttaaag tcaacagact aataggtagc atagctgaga 8700
 ttgttgcacta ggcatttta ctccctgc gcaagagtct cttggcattc ttgttgcacta 8760
 gcatatttct taacctcaact gaggctact ttccctttat ataataatggg gttaaagagcc 8820
 ctccaccctgc ctggcacaca ctggtagtgc tcaacat tggagggtgt tagtttaaag 8880
 gcttcatgga ctctataatgt tcaacaaaag tgctgttaac ttctttctgg gtctcaggct 8940
 cctgtatgttgc agtcactgttgc gcaaccctgc catctgtgttgc ttagtgcact 9000
 cacacttact aacctaaacc ttgttgcact gctgtggct tctccagaa gtttttactc 9060
 atttgcacttgc ttatctttt agggaaacagc cagcccgtag atcattaaagg ctggctatttgc 9120
 gacagggggc tggggcctgc ctgacagagg aaggaaggc agacatctgg ttcttcctct 9180
 gcccctacaa gagactccag cctgaccaca gagtggtact cctaggatgt agcagcagca 9240
 tatgagcttgc aatgtgcctt aatcctgtc ttactttgtat gaaagagagaa ctaaggaccc 9300
 acagatgttgc tcaatcttca taataaccctt gcaaccccca ccttttttgc agaacagggt cttgtctgt 9360
 gagatagata actgatataa attaaacgtt gtatataaagaa cctcacttgc attatctgtat 9420
 tcaatcttca taataaccctt gcaaccccca ccttttttgc agaacagggt cttgtctgt 9480
 tggccaggctt acagtcact ggtacaatca tagtgcacttgc tggacttgcact 9540
 caagcaatccccc tcccccacccca gcttgcacttgc cagttggac tacaggcgctg ccaccacacc 9600
 ttggccatttt tttttttttt aagtagaaac aaggctttat taatactatgt ttggccaggc 9660
 tggcttttgc gtccttgc gtccttgc gtccttgc gtccttgc gtccttgc 9720
 gtaaggccact gtccttgc gtccttgc gtccttgc gtccttgc gtccttgc 9780
 aagggttttgc gtaacttgc tgggttgc cagatgtat tggacttgc tggacttgc 9840
 ctcccaagag agtctgttttccacttgc tcccttgc gtccttgc gtccttgc gtccttgc 9900
 agacttggag acagaaaatc tgatgttgc tggacttgc tggacttgc tggacttgc 9960
 tggccaggctt ctttgccttgc tggacttgc tggacttgc tggacttgc tggacttgc 10020
 aatccatctc atggagatgt tggatgtatc aaggactctg aaaaatcttgc aatggttttaa 10080
 tggacttgc tggacttgc tggacttgc tggacttgc tggacttgc tggacttgc 10140
 tggacttgc tggacttgc tggacttgc tggacttgc tggacttgc tggacttgc 10200

-continued

gttggcatac tagcctggcc tcttaattaa ttcatattaat agcttattta tttttgagat 10260
 aggttctgct ctattgccc ggctggagt cagtggcatg atgatagctt actatagct 10320
 caatctccca ggcttaaaca atcctctga gtagctggga ctacaggcac acactaccat 10380
 gcccagtaa tttttttta atttttgtt gagacagggg cttgcctgt tgeccaggct 10440
 ggtctcaaac tcctgggctc gagatctcc cacctggcc tcacaaagtg ttgggattac 10500
 aggtatgagc cacggcacct ggcctggct cttaaactggt tccctaagac agctgaaat 10560
 agagaatgtc atggagcatt cctaaaccatg ggctccagcc tggctttcat tctgtttctc 10620
 ccctgaaaca acattccctt agtaatattc cgaataacag cttcatcagt ctgtctaccg 10680
 accacttcc aggtttcata ttatatgacc tcccaaactg cactaagggt tgtatttagag 10740
 aaaaagtggat aaagttcggg gtcaggctgc ttgagcttaa atgccagctt cacttaccag 10800
 ccacctgacc atgagtcagc tgcttaacca ttctttgcca cagtttcctt gtctatgaaa 10860
 agggaaatgg ctcccacctc aaaaagtgtt taacattaaa ttcaatcatg tattcaaaagt 10920
 cctgagcaga atgtctggcc atgactggga cttaaacagat gtttagcattt attatttagt 10980
 tctgtcagtc ttgaaatgtt ctctccctt ggctttcatg acattccaca ctctctgg 11040
 tttctcttac ctctctggta atacctgtt gcttacccctt ctttgcctag ctctggatg 11100
 ttaccatcc ttcaggcgtg ctgtttctc cttaggcagt cttacacaca ctcatgactt 11160
 ccttccatttgc tccctccacac actgatgacc cttaaatcag tatctccagc cttaacccctt 11220
 ccactgagtt cttagaccat atgttactat atcaacctgg cttgtccatt tgaatgttt 11280
 ccaggcactt cagactctct tctcttagact ttgtctggact ttcaactctc cccctaaaac 11340
 tggctctct tccactgaaa catgtatgtc attgagaggc accaccatcc acccagtgcc 11400
 taagccagaa accttaggaat ccttgatacc tggttctctc catcctgcat atccaagct 11460
 atcagtttta tctctaaattt atattttgtt aggtttactt ctttccctttt ctccaccac 11520
 caccctgttc caagctacca tcatctcacc tggatgtctg caatagctc atctccacac 11580
 gccaactctgc accccctaat ctgttctcta tagagcattt ggaaggagtg atttttgtt 11640
 tttgttttgtt tttgttttagt acagactctc actctgttcc ccaaggctgg agtgcagtgg 11700
 cacaatttgc gtcactgtca acttctgcct cccgggttta agcaatttctc ctgcctcagc 11760
 ctccccaaatgtt gctgggatataaggcaccgc ccccatacc acgttaattttt tataattttt 11820
 gtagagatgg ggttttggca tggccaaactt gctagtcg aactccctgac ctcaagtgt 11880
 ccacctgttc cggccctccaa aagtgtctggg attacaggtg tgagccactg cacctggctg 11940
 gaaggagtga tcttaaaaaaa aaaaaaaaca aaaaaaaact tgactgtgtc actctgtgtt 12000
 gtctctcttatac cttccacaact tcccagtgtt cttggataaa gacaaaaatc 12060
 cttaacttgg ccaggcgcgg tggctcacac statcatctc agcaattttt gaggccgagg 12120
 caggcagatc atgaagtcaaa gagattgaga ccatctggc caacatggta aaacccatc 12180
 tctactaaaa atacaaaaaat tagctggcg tggtggcggtg tgcctgtactt cccagctact 12240
 tgggaggctg aggaggaga atcacttga cctggggaggc agagggttgc gtagccag 12300
 atcacgcccac tgcactccag cctggctgaca gatgttactt ccatctcaaa aaaaaaaaaa 12360
 aaaaaaaaaa ttcccttaattt tggccctacag tagagccctc cgtaatgtgg cctctcttca 12420
 catctccacca acctccctgtt cccctgactt cagccctacc tctctctgg acaggccctc 12480
 cttctgacaa gggctttgtt cattctgttc cctctgtctt gatgtcccc ttactctgtt 12540
 cacttaactc ctgccttatacg ttttagatctt tacctggatg gtcagagaa atataagaat 12600

-continued

aattcctcac cctgaaaaat aggttaggtc cctgttttat gtttcatag accttcctt 12660
 tgaggctttt ttaaaaaaag tagtttaat ctcacattta ttcatgtat catctccta 12720
 atgatatctt aagacctcta atagaacaat ttggcatgg actgtgggt tttggccct 12780
 cattgtgtca gcactgagca tattgtggc ataggagga tattgttga atgaattgt 12840
 agaggtggcc aagagatatg atgtaaatgtca ggctttccc tgccctccc cttcccttc 12900
 cccacatctt tcctatagca gccaccgtgg ctgcagttac tgtaaatggc aagacggaa 12960
 cagttccggc cattgggtt ttttagaaaa ttgcctgca gtgtcagggt gataagttaa 13020
 agctttgtct tttgcctcta gaggagctat cccatagtga gtagaagcca gagaagctga 13080
 ccccaaggagt ccttcttcc agcagcaggat cttgagctgc acttctctgt agtacaaatc 13140
 caggcagggaa caagccctag gtacccccc agaggagggc aagagaggaa gaatgagttc 13200
 agctactcta gcccaccaaaat tgattatgaa ttgcctgaa atctgaaaaa tttcaattcc 13260
 aatcgtaaatg ttgtttttgt tcattttgtt ttcttaattt gtatatttga aagatggcat 13320
 taactaaaga tatataattca atatagatgt gaaaaaatgg aataacttgca tagtatctt 13380
 tacttataagg tgattttatgatg tggggagtttgg ggtggatagg ttggcagttc ccccaagaag 13440
 ttggaaatgaa agtttgcct ctgtgagtttgg aactaatttgg atccacaagt aatgaaagca 13500
 gtattgtgtt gtagttaaga gcacactcta gaaccagatt gcttagttc aaatcctgg 13560
 tctgcctttt attatctgtg tactttggc aagttacttg ccctttgtgt gcttcatttt 13620
 ttcacatcttag aaaaatggaga ggcaggcgt agtggctcat gcctataatcc ccaagacttt 13680
 gggaggccgcgaa ggccggcaga tcacctgagg tgagaagttc aagaccagcc tggccaaat 13740
 ggtgaaaccc tgcgtctaca aaaaatccaaat aattagccag gcatgtatggc ggggtgcctgt 13800
 aatcccaagct acccaggagc ctgaggccggg agaaacactt gaaacctggaa ggcaggagtt 13860
 gtagtgagcc aggattgcac cactgcactc cagcttgggtt gacaagagct agactcagtc 13920
 taaaaaaaaaaa aaaaaaaac aactggaga tacaggctgg gtcagggtt tacacttata 13980
 atatcagcac ttggggaggc ctaggccgggaa ggattgttttgg aactcaggag tttcaagatc 14040
 agtctgggtt acaagacaaatg acctcatccc cacaaaaat caaaaatataa gccaggcatg 14100
 gtggctcatg cctgtggcc cagctactca ggaggctgag ggcaggatgt tgcttgagcc 14160
 caggagggttggcaggctgacatgacatgactcaatggcactatcatgatggc tggatgacag 14220
 agcaagaccc tatctcaaaa aaaaaaaaaaa aagaaaacgca gccaggccgc tttgttcacg 14280
 ccagtaatcc cagcacttggc ggaggccaaag gcaagggttggcacttggatgttggatgt 14340
 agactagctt ggcacacatg gtgaaacccccc atctcaactg aaaaatccaaat aattagccag 14400
 gcatgttggc atgcttcctgtt agtcccacatg actcaacttggcaggctgacatgac 14460
 cttgaacccca ggaggccggg gttgcagtttgg gccaacatca tgcactgca ctccagccctg 14520
 ggagacagag cgagactctg tctcaataaaa taaaataaaaaa taaaataaaaaa 14580
 taaaataaaaaa taaaataaaaaa tggaggccag caggcacgggt ggctcacgca tgtaatccca 14640
 gcaactttggg aggccgggggg gggccggatca caaggtcagg agatcgagac cttccctggct 14700
 aacacagtga aaccgcgtctt ctactaaaaa tacacaaaaat tagccaggca tggtggcagg 14760
 cacctgtatgtt ccctgtactt caggaggctg aggccaggaga atggcgtgaa cccggggaggc 14820
 ggagcttgca gtgagctgag atcgcgcacac tgcagttccag cctggccgcag agagcaagac 14880
 tctgtctcaaa aaaaaaaaaaaa aaaaatggag gttggggccgcg gttggctcgcc cctgttaatcc 14940

-continued

cagcactttggaggtcgag gccccggat cacctgaggc caggagtcc agaccagct 15000
 ggccaaatgtgtaaaccccttgtctactaaatataaaaatattagccag gcacgatggc 15060
 aggcacccatgtatcccaactttaggaga ctaaggcagg agaatacgaaacctgg 15120
 gatggagggttgcactgtgtcttgcactgtcccttcactggatgtcaggatgtc 15180
 tcaaaaaaaaaaaaaaaaagaaatggaga tacaacttacttaccccttaccaacc 15240
 taccctcaca gtattactgttgaataaaatgtgtgttagca ctggaaacac tattcacaga 15300
 gcactcatgtatgttttttccttgcactgtatgttgcactgtatgttgc 15360
 tgaataatgtgtgtatgttgcactatgtatgttgcactgtatgttgc 15420
 aataaaaactttagttaaaaaccatataaaggccgcgcactgtatgttgc 15480
 ccagcacccggggggccggatgtggcaga tcatttgcggatgttgcactgtatgttgc 15540
 tggccaaatgtgtaaaccccttgcactgtatgttgcactgtatgttgc 15600
 aggcatgtatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 15660
 ttgaacccatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 15720
 gtgacagaggatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 15780
 taagtggccatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 15840
 cagataaaatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 15900
 gcttagtttgcactgtatgttgcactgtatgttgcactgtatgttgc 15960
 ctttattttatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16020
 tacttaataaattttatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16080
 tttaatagtcatttataatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16140
 cagaatgttgcactgtatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16200
 ggccttagattttatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16260
 caccagtcaatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16320
 gtaatgggtatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16380
 agcccgagatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16440
 gatttttatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16500
 ttgagatattttatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16560
 gcatatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16620
 atgattatatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16680
 tatcagctgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16740
 ctgttagtgcactgtatgttgcactgtatgttgcactgtatgttgc 16800
 gacttttaggcatttataatccatttataatccatttataatccatttataatcc 16860
 ttggcattttatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16920
 ctcttcactgttgcactgtatgttgcactgtatgttgcactgtatgttgc 16980
 ctactctcttccacccatgttgcactgtatgttgcactgtatgttgc 17040
 gagtttacatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 17100
 ttgcacatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 17160
 agtctgccttgcactgtatgttgcactgtatgttgcactgtatgttgc 17220
 ctccccaccaatccatgttgcactgtatgttgcactgtatgttgc 17280
 ctcaatgttgcactgtatgttgcactgtatgttgcactgtatgttgc 17340

-continued

ttctccaagt gtggagccgt atctagttt tctttgtatc ccagagctt gcaaaatgcc 17400
 tagaatgtag tgggtgtca gagtgtttgc tgggtgaatg atgtatttgc tgaacgactc 17460
 tttggacact tgaataaaatg ccatccagta tgcaccattt ccattcttc getctacaat 17520
 attcttttag gcaagagctt atcttttagt ggtataagat aagctcaaaac ttatgttagac 17580
 taagacccca gtcgttaat gtcattccca agtcttaaac catcaaaacc agggccctcaa 17640
 ggaatggcat gccttctgc actgttagca cctgtgtgc ttatgttagcc gttttttca 17700
 tttttccccca aaaaagctaga gtcccttctc ccattggcag tgctgaaatg gtgctaaca 17760
 attcttctc catactgtttt acgattacaa aaaaacccctt cagcatctca tgccagactt 17820
 gagtttaagg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 17880
 ccctatagag tttttgtca gatcagggg tgctccactg ccattcattt cactgactct 17940
 tgcagaagca ccgtttctca agttggctaa tttttttttt tttttttttt tttttttttt 18000
 ttgttttagt tccagagata gcaacttcat gaaatgacgc tatctttagt aatcactttt 18060
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 18120
 cagtcactg caatctccac ctccgggtt caagtgttcc cctgttccca gctcccgag 18180
 gagctgttac tacaggcgtca cacccttactt cctggctaat tttatgtttt tttatgtttt 18240
 cggggtttca ccgtgttgc caggatgttcc tttttttttt tttttttttt tttttttttt 18300
 tcagcctccc aaatgtgttcc gattacaggtt gttttttttt tttttttttt tttttttttt 18360
 tttttatacc ataacgttac caccactgtcc gcttgcacaa gttttttttt tttttttttt 18420
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 18480
 gcttaacaaa atgtcaatca ccaaaatattt atttattttt tttttttttt tttttttttt 18540
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 18600
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 18660
 gaaaaagaaa agcattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 18720
 aaaaatgcaga ttcttagccca cagtctcagc gattcttgc tttttttttt tttttttttt tttttttttt 18780
 tcaggaatct tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 18840
 acacttttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 18900
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 18960
 tagacccctgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19020
 ctggtcaccc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19080
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19140
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19200
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19260
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19320
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19380
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19440
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19500
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19560
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19620
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 19680

-continued

-continued

ctctccagct gggcagccct ttcagtatcc cgtatgttat ttccccactt ccagccacc 22140
 tcacccctc tggccctt gtgtgtcccc tcggcttaga tcctgaccc tcgtcaaga 22200
 gtttaaactc aacttgagac ccaaggaaaa tagagagcc tctgcaacct catagggtg 22260
 aaaaatgtt atgctggag ctattttagag acctaaccas gggccagaca gagagagtga 22320
 cttgctaaag gccacatagc tagcccacag tagttgtaaac aatagtttta atgatattaa 22380
 tggctaacat ttatcaacct ttaatgttc ccagacttg tgccaaaggc ttacatcag 22440
 tgcattgtcg cattcaaaacc cagacagtct ggctctggc ccaggctgag ctttggata 22500
 gcatggtaga acgttgcata taatgtctag tctgggtca aatcctggct tcacttctca 22560
 catttacagc tgagtgcact caggcaagtg atttaacctc cctgtaccc agttgttta 22620
 tctgtaaaga gaaaaatcac agcaactgtgg aatagtgggg gttaaaattc attcatacaa 22680
 gtagtgcgtc aagcaatgtt taatacaggg tgagcacctg ttcaagtgcctt ctttcttcg 22740
 gctgcctctg gggctaggt gtgggtctt cgtggatag atagatagat atggctgagc 22800
 tctgcacaaa caccacagc tggtttcac tattagaggt agttaacaca gtgggtgagc 22860
 tctgtgggtc tagaacagag gcccggcaagc tatggccat tgcctatattt aatacggct 22920
 gtgattgatt gatttttttt ttcttttta gacagagttt cactttgtt gcccaggctg 22980
 gaatgcaatg gcacgaaactc agtcacccgc aacctctgc tcctgggtc aagcgattct 23040
 cctgtctcag cctctcgagt agctgggatt acaggcatgt gcccacacgc ctggcttaatt 23100
 tttgtattt tagtagagac agggtttctc catgtggc aggctagttc cgaacttcca 23160
 acctcagggtg atctgcccgc ctcagccctc caaagtgtg ggattacagg cgtgagccac 23220
 catgactggc ctgattgact gatttttta gttagagatag ggtctgggtt tggttccag 23280
 gctggctcata aacttctggc ttcaagcagt cctccctcct tggcctctcg aatgctggga 23340
 ttataggcat gagccactat gcctggccata tatgacctgt gattttaat ggttagggga 23400
 aaaaaagcaa aagaatgtt tggatgtt ggaattaca tggaaactcaa atatacgatgt 23460
 cccagcctgg gcaacaaagt gagaccctgt ctctacaaaa aataaaaaaaa aataagccag 23520
 gggccggccgc agtggctcac acctataatc tcagcacttt gggaggccga ggcaagtgg 23580
 tcacccctgagg tcaggagttc aagaccaggc tggccaaatat ggtgaaaccc tggctgtact 23640
 aaaaacacaaa aaattagccg agcatgggg catgcgcctg tagtcccacg tacttgggag 23700
 gctgagacaa gagaattgtt tggaaacttggg aggccggaggt tggctgttgc caagatcg 23760
 acactacact gcagcctggg caacagagcg agactccgc acacgcacgc acgcacacac 23820
 acacacacac acacacacac acgctgggtt tggccggccag cacgtgtggt cccaggatgc 23880
 actggaggct taggttaggag gatcacttgc gtttaggtgg tggagactac aatgaaccat 23940
 gtttataccca ctgcacttca gcccggccaa cagtggtgaga ctgaatctca aaaaagaaaa 24000
 aaaaaaaaaga aaaaatctt tccataagta aatatctgtt ggaacatagc catgtccctt 24060
 agtttatgtt ttatatatgg ctgttttgc cctataatgc cacaattgag tggccacac 24120
 agtctgtatg gcctgcagag cctaagatat ttgcctctcg gccccttaca gaaaaagtgc 24180
 cttgacccctgt gctcttagacg catatgtacc aggtttgaaa ctcagccctca cagctgggtg 24240
 tgatggcagc catctgttagt cccagctact ctggaggctg aggtgagagg atcacttgc 24300
 tccagaaggt cgaggtcaag attgttagtga gcccattgttgc catcaccgcac ctccagccctg 24360
 agtgcacagag agagaccctg actcaaaaaa aaaaaaaaacaa aaaaaaaaacccaccctcacc 24420

-continued

acttatcagc tattttgtctt gagaatagtg acataacccc tcagaaccta tttctaatac 24480
 tggtaaatga ggctgtatgac gtttccctct tttactggca atttaaacat gatggataat 24540
 aatgcataag cacttaaacac agggccctaga agatattaac tgctcaataa atggtagctt 24600
 cttaaacatgta ttcaaaaccca tttgtcttta tcacatgtat tttttatcg actttccctgc 24660
 ggttggatgg gaaaaggctc ctttgcataacc ccatctatcca tttttatcg actttccctgc 24720
 catggttcac agtaagagat agaagctgca cggtgacttc tggctcttta caatggtag 24780
 cgggtgtgtgc ctggtaaggg agagctgtatg tcaactgcacc aatccatgtatg 24840
 agtggttctgg tttccctccag cagccctgtat ttttcccttta caatctgca ggcaggggaga 24900
 caaggggctttt ctacatggta ggctctgggtt tggtcatacgat cacaactggg ggctgttcag 24960
 gtggggctccc attccagata ccttaggttca tcaatccctt ttggcaccacc accgccttttt 25020
 ctccctcatg ccccatttttt cagtttggaa agcatggta tcacaggaca agtagaaagaa 25080
 gctccactgt ccactgaggc caatggatgg ttttctgtat gtaacacactc agtgaatag 25140
 gagtgaatga gagtaacactg ggctccatcc tatttgcaga gagctttggaa aaagattttt 25200
 ctcccttaaag agccagaatg aagccctggta gtggggagagc tccagctcta gatgcacatg 25260
 agccctacatt taaaattccag ccctggccact gactccctttt ttgacccttga gtgaggattacc 25320
 taatctctct gtacccatct tttcttgcata gtagagttggg aataattccct gtctcagaga 25380
 aataaaaagag tgcataatagt gtttgcacaca tggagacaca tcaggtgtat gttataactc 25440
 tggggctctgt ttccttattt gcaacacacgc cctggccctgg agtggaaatgtg gcaacctccca 25500
 ttgggtcagct cttggaggctg tcccccaggac aggcagaggg agggaaatgaa tggggccct 25560
 agtgcacccaga cagaacagat ggcagctcag agcttaggatg gctctcttgc cctgtctctc 25620
 ctaccagagg tccccccgc tgggtgtgtat cttccctggac ctggccatctt ctgttttttt 25680
 tttttttccca cctccaaagca gaattactgt cctgttaggca gtcctctgtc ttgaggacat 25740
 ctggggccacg atatgttccac actctatccct gccttgcctt tccctgagct caggatggac 25800
 gctcaattgg tcccaggat tgcctgcacgc gctgcctgc acgcctcgatc cagcccaagct 25860
 ccaccccttg cctgcacaggat ctgtttccata acagctgtatc caaccacaca cctcggttct 25920
 gcggggagccc ctcctcttcc tccctccctc cctcatttag ggggtggact gaagaagaag 25980
 gcttaacttgc cagcagctat ttttttttttgc ttagtccaccg gcccctgtatc aagaatgcac 26040
 gtgtgtgtgt agcctccacca gagaggtcgat tttctcgagat tccagagggg ccgcctgagc 26100
 ttctgagaaac tagggaggag ccatcccacgc catgagcccc tggggaaatc tgctggggc 26160
 caatggccctt gtagtccatca ggctcccgca gctgctccgg agggagaggt gagctcagg 26220
 cagccctgcctt gcaaggccatc gtagccggag ccccgccctt gtcattgggg ccatctacag 26280
 cccggccctggag gcaatgcacag acggatttgc agctgagccct gtcattatgg tggggaaaga 26340
 agatggggag ttacttgcac gtcctggctt acttcacccatc cagagacatg tttcggttag 26400
 ttgggtctcg agtccctctc tccatctctc ctggccctgt gtcctgagat gaggggtggc 26460
 tccctaaatc tccctctcac tttagtccctt accatcgat tgcctggggca gaagccacgc 26520
 gaggttatac ccaaggagaa tggccctgtat gaggatcccc cattatgtcc tggaaatgtt 26580
 gagggggaggg atataacccatc aagggacttc ttagggagctt ccagctcccc ttctatccca 26640
 gacaaacccatc aaggagccctc caaaatgcacgc cactgacatgc cccattgtatc atgttactgc 26700
 ttccggggggg aatagcccaa atagatgtat gttccatgtat ctcacatgtc ttacctgcgg 26760
 gccatgtatc cttggccatggaa atttgtccca acaaggagggat tggggcaggat tggccaaact 26820

-continued

gtggaaactg gcaagtctcg ggtgtgggta gcctggata cagtaggcac cttataaacg 26880
 tttgttctct taatggcagg cacattgcc tctggccttg aagggttct gagctccag 26940
 gtgaatgtat ttgcgtgggaa aagacctggg cgagtgcctc taagactgga gcaatggct 27000
 ttagagtgtt cctgagctgc tggccagcc cccacaccc tcagtcctt agggcttaatg 27060
 acctccacga gcctctctt gtggggcttc tcagaggag atgtggaaac tctacccctta 27120
 acctggcttt ctgtgtcat tgcggccatc caccctccat agaaactccc cagggggtt 27180
 ctggccctctt gggcccttc tgaatgggc cattccaggc tagggggggg tttgtttca 27240
 ttctttgggaa gcaagctgtt gttccaaaaa ggctgcctcc cccctcaccag tggctctgtt 27300
 cgacttttcc ctgtgtcat ctgtgtcat ggtccagtc ccagatcttgc tggccggat 27360
 actagtcagg tggccaggcc ctggggcagaa aagcgtgtt ccattgtggg ttgtggaaatg 27420
 accggacccct ggttagattgc tggggaaatgtt ctggacaggg ggaaggggggg aggaaactgg 27480
 tcctcaatgc tgactctacc aagcgcctcg ctagacactt tattccctttaa tctctcaaca 27540
 gcctaaagat attatataatc cccattttac agatggggca accatgttca acagagttaa 27600
 catatggggc ctcaactgggc agcttttctt gtcttcctga ctgtgtctca tccttcagg 27660
 ggctgcagggt ttgtttctt ctcttaggg agaggaaattt ctcagggttgc ttttcctctc 27720
 cttagcagaga gtaaaaaaaag ggatagtttgc cctgactgtt tgaagggttg gctgagattt 27780
 ttttctaaag agccaaatggaa aattgtatctt gagtttagga gaaagctttt acatgtggaa 27840
 ttaagatgcc aagtgttggaa gtagccacat ttcaaggctt cattaaatcc tcttaatctt 27900
 gggaaaggcag cttagggagaa ggggtgttcc tttagggagcc aggaactata ccccttttac 27960
 ccttggagag gcaggggaaagc caggggaggac acaacttctc aggaagaggaa gaagcttagag 28020
 cagatagtgaa actctcaacc tgaaccttta agggccagac cactaatgcc acccaagttc 28080
 acctggcggtt tggcttttgc tggccaggc ttctggaga acctgtatctt ctgtccctca 28140
 ccccaagct ccgtttggcc agctagatgc tgggggttgc tgactgttgc tggtagacat 28200
 tcttccttc cccaaataag agggccacatt cctgaagtca ctgtgttgc gatagctgcc 28260
 acacagggtt ctttttttttcc agggaggggac cacccagacc ctgtgtctc ccaggatcc 28320
 gttaccacat cactaccccttgc tcaagaaatgtt gtttctggca tttagcccttc ccttttttgc 28380
 tataggatatt cctcaaggcc ttctttttgg gcctcaggat tatttttttttgc agaaagtttgc 28440
 agcttagactt cttggggctcc tgaacagggtt ctttgcgttgc ttctgtggaa caaattaaatgtt 28500
 tcttgacccctt aggcctctgg gggaggatcaaa agtctatggg agtttctgggg ctgtgggtgc 28560
 aaggaaatgtt acgcaaccatggtt attttttttttgc gatagcttgc ggcgtgacat gtgagggagg 28620
 aagaggggggc aaggaaatggaa agaataacaac ttctgtgtcc catacacccttcc tgcctgacag 28680
 gcccatacata ctcagcagag aatgcactgtt ctttcctacc acactagcgtt gaggaggatgc 28740
 ctgcatttttttgc cactgtgttgc ttcttcatttgc aatttttttttgc ttttttttttgc 28800
 gtttttttttgc gtttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 28860
 cacttaatgtt ccaataactgtt ttttttttttgc ttttttttttgc ttttttttttgc 28920
 caaatgttgc ttttttttttgc aatgttgc ttttttttttgc ttttttttttgc ttttttttttgc 28980
 ctggcccttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 29040
 atctgttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 29100
 tcctcagaggc cacacaatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 29160

-continued

actgcagtagc aactacacaa aataacttggaa actacagttct tccctgggttt tgggtggaaac 29220
 tgaatcagtg cactctagca acacttattt cttgctgttc gtaggcttca ttatgtgttt 29280
 ggttaatttt taaaacaaac aataacatata tccataataa ttacagctta atggcagac 29340
 tggttcagtc tataggatct gcaggaagga ggagtaataa agggatttt gactgagctc 29400
 ttatggaaaca gagtctctct agggccctgt catatctgcc cttctgggcc ctggggaaaa 29460
 gttggcatcc ccagttgtgg tgcctccag gtgcctcag gctgtgtgg agggagctc 29520
 ccattctctc cttcagccca ctcaattcag aggcttaggg ctggaaagaag ctctcttaca 29580
 actggctgtt cactgggagg ttaagggatg accatccagc caggcctcc tcaggacatg 29640
 ggagggctta tgcttaaca tggtaatac cactgcaata atgactggtt ctttacccc 29700
 ataagggttga gaatttacct gtaaacattt ttgtctgaag aatttggatg taagtgaggg 29760
 ctgggcctct atcttatctc acttggcttc ttcagcaca gcaccttgc tgcttgcct 29820
 tacacatcct agatgcacag taacttattt ctaattattt gaaatctatt agaatcaatt 29880
 gatttcagct gggcttggg gtccttcct gtaatccag cactttgggaa ggctaaaggct 29940
 ggaggatcac ctgagtcacag gatgtttaaga ccagcctggg caacatagg agaccctgtc 30000
 tctacaaaaa ataaaaaatt agccaggat ggtgggtgtgc acctgtatgc ccagctactc 30060
 aggaggctga ggcaggagga tctcttgagc ctgggaggtc agactacagt gagcaatgt 30120
 tggccactg cactccagcc tgggtgacag agtaagactc tgccttctaa aaaaaaaaaa 30180
 aaaaaagttt attttatattt ggtatagataa ataattcatt ttaggacctt ttttttac 30240
 ttacagaaat ctgttttattt ctgggctgag aagcaggatc atattgtatg gcataggaga 30300
 aaaaagggtc tgcctgcatt tgcccttggt ggtctcaat tggggagggaa aaaaaatgaa 30360
 cacttactgg ctaccctctg tgagccagc atcatgcaag acatctgtac ataatttaat 30420
 tctcataacc ccataagata ttattagcaaa tgcataatgtt agggaaactgtt ggctcagat 30480
 catgaagtaa ctggcccttgg gtgacacaga tggtaaatgg cagagaagga atatggatcc 30540
 aggtcttggaa agaaaaatc tcaactgatt atcttttttta aaaaactcat atgttcttg 30600
 ctgactcaaa aggtctctgt gtggatctgg gttgaccac tgaactgacc atcagggttc 30660
 catgactttt gatatgtccc aagccctcag aacccttcag taatgttttgg gaagatgat 30720
 tttggggatgt gtccttaggc atagcctcag cgtatgttgg cctcttaggtt atctccctca 30780
 acctgaggat ttcaactctgg ctccctcagga cagttggatg actgggttcag 30840
 acctcagttt taccacctcc cagctggta ctcttctacc tacagccagg gcagattttgg 30900
 actttcactt gaaacttcca aaaattgaaa ggtagaaaaa cagccttggc tttgggaaga 30960
 acgtatgtatg tccatggccat ctaagcatatc gaggtgggac atgttgcagt agcaccttac 31020
 agttccaaag tgcgtttctgg gttctttttttaaaaagaaca gagactgttgg gggaaattgaa 31080
 cactgttgaag tataatgttggtggtggtggtggtggtggtggtggtggtggtggtggtgg 31140
 ggagaagcat cacgaagtgtt taacactcaa agggcttgc gctgtcagggttcccttgc 31200
 ctatattttca caggttgagaa tcctgaggatc cagctgttgc gatgtgtgttccactccgg 31260
 tgacatgttgcagatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 31320
 catcaattat gtatggca gctttgcaga atgatttgc tttactctgttcccttgc 31380
 ttctgtaaaaa cagggataat cctgctaccg taggggtgtc aggatgttggtggatgttgg 31440
 taaggatctt catataggac ctggattatg gctggcatcc aataatataatgttggatgttgg 31500
 tgatagctaa gcttagactc tgaagtctacatc catggcaacttcttaatgttggatgttgg 31560

-continued

cagttgtgtt ctgtggcaaa acacagctta gggatccata cccagccctc ctgtcagctg 31620
 ttcacccccc agttcttcag agacatgtgt ggcagtact ttggccacat agctggctgt 31680
 gccccttaaa ggccatccctt gacacagata tgtggactgg tgacgttgc ctccagccag 31740
 gtgttcttcc cagcagctgt gcctggctgt ctccctgcatt cctgtacttg tttgtctccc 31800
 tgctccctctt cctggccctg gccagagcta cttgcagcaa aaaaaagcag gatattggca 31860
 atggaaagga gggtgtgttc tgggtctccc atgcctgcg gcgacacatac cattgcaagg 31920
 gcgtaacaga gcccaggcct gcattttgggt gcaaataagt ctgcacacag aaaaaaagaa 31980
 ggacacttggc accaggagcc atggaaacctt tgggtctccc tacctgggct actgggtctt 32040
 gccacttcttta ccattttcag tttggaaata tttgttaagg ctttgcattt ccagggtctt 32100
 tgcttggtgc tgagtctacc aagagtaagt gggatgtgt ttttgcctc agggagctaa 32160
 cagtcttagt aagaagaaag atgggttgcgg aggaacttct aagtcaagag gcaaggaggca 32220
 agaaggaagc ccctgtctt actgcctggc ctctgttggg cacccatag ttcttcagaa 32280
 ccacatttaa tcctcactgc agggcaggca tagtggctca cacctgttaat cgccggactt 32340
 cgggaggcca aggcggccag atcactttagt gtcgggagtt cgagaccgc ctcaccaaca 32400
 tggggaaacc ccgtctctac taaaaataga aaaaattagcc ggggtgtggt gcatgcgcc 32460
 gtaatcccaag ctactcagga ggctgagggt ggaaaatcactt tggactcgg gaagcagagg 32520
 ttgcagttagt ccggatgtgtt gccactgcac tccagctgg ggcataagag caaaattcca 32580
 tctcaaaaaa aaaaaaaaat aacttgcactt accttggaaag taggtgtatga 32640
 cattggcatt tcacaaatga gaagtgaagg ggcttagccca agatcactta ggtggtaaat 32700
 ggtggtgcta agattagaac ctcagatcat cttagggaaa acacagatat gcacagatgt 32760
 aagggggaccc agggattttgt ttgttcttctt gtttccacagg tggggaaaca acccagagag 32820
 gggaaaggggc ttgtcccaagg caatttagca cccaaagaact tgaacccata tctctcttctt 32880
 cctcatttttag agtcatccc acatgtatct tatattgaga ggagtgttag ccacatacca 32940
 agaacagttt tccctctgc ctccaaacctc actgtgcactt tttgagacac ttccacagcca 33000
 tactttcat gccataccca gcccctaaga ccctgaattt ccccttccat aagacaagta 33060
 ggaaaagcta tagggtaaaa atagccatca gtgtttgttgg agcaccagg aggaatttgg 33120
 cactccagaa agataaaaggg atttcctcaggg acttgcatttctt cttagacttcc ctgcactc 33180
 tgcttcaactt cattcctggc cttttttctt acctcccgca gtgcactgaa gtagttagaac 33240
 tcactgtggc ctctcaccctt gcatgttga gttttatata gactttcttctt tccactactc 33300
 ttcataagctt catgaaagggtt gaaatgtgggtt gcccgtgttata tttatctttt atatctgcag 33360
 tgcttagcaa gttataataa tgcacttgcc tggcaaaaagg ctttctctca tacattagct 33420
 tattttcttctt tcacatttggc tttttttttt aataggatgc tatttagttat tttcaatgag 33480
 agaaagctac taagagaagt tgcacttgc tggacagttt gttttttttt aatgtggactg 33540
 ccattacattt gtcactatctt ttaatagaag ttaacacata ctgagtttctt actatattgg 33600
 gtctttttttt tttttttttt tttttttttt gggggat tttttttttt gttttttttt tttttttttt 33660
 gggaaacggcactt ggtgcattttt tgggttccatca caacccccc ttccctgggtt caagcgttcc 33720
 ttctggctca gcccggcttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 33780
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 33840
 gaccccttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 33900

-continued

-continued

aggacccca aggagagtgt tacaggctga gaggaggata tacttaggtt gtcttttaggg 36360
 aatcagaaaa ggagactctg gaatagctg gcagagagag gggctaccc tcataacctc 36420
 tctggacaaa cgacttaag catagtgaca gatattgcca ccctgtattt gaagaactga 36480
 tcttttttag tgggatgt tacttctggg gatattctt cataactgag accaaaacag 36540
 ttttgcag tctcagaaat gacaggaggt accaactgta cacttcctt ggaagctcta 36600
 gggcagagag tgaaagatg gatattgacg gggccctgc ttggaggta ttcaccacc 36660
 cctgtctca ctccagcaac agtgataact cacttccttc ctccctttgt acacccttct 36720
 cccccacctgc tcacaggctgg ctggggatgt caagtaccac ccagagtgt ttgcctgtat 36780
 gagctgcaag gtgtatgttggg aggatgggg tgcatatgca ctgggtcagc atgcccaccc 36840
 ctactggtaa gatagtggc ctttgcatac cctctccat ataagagtgg ctggcgggg 36900
 gggacagtgg cagggtgagt tggcagaaag gatgtttagg gtatgtcagc cattggattc 36960
 ttaccacagc agtgcttta accagctttttaacttggtaa gcaaatgtt ttacacatgt 37020
 ctctaccctt tttcccttacc aaccttggaa atgtcttcac tctgcctgc aatcctccca 37080
 gtgggaggca ctcttcaagg acgatccag aacattaaag tcaaagaccc cttagagctc 37140
 accctgtcca accacccttgg ttgataaaag aagtccgct gggccctatgtt gaaatgaaata 37200
 gtacaaggc aaggttctca ttgtgagtc aaggttagatgtt gaaatgaaacc cagaccatct 37260
 caccaccaacc cagggcagtg ttttccaaa tataccactt gctgcagatc tagctcagca 37320
 cccccagttcc cagcccaccc tgagaaccca ggctctcat tctgagcagc cagcttagat 37380
 catgacaaag agggtggttag tgagactatg ggtactgtt cttaaagcca catgggtcag 37440
 tggttgttgg gggcccttctg tggggactc tagcatcttta ttcccccctg tggccctctcc 37500
 ccagtggaa gtgcacaaat gaggtggtgc tggcacccat gtttggaga ctctccacag 37560
 agtctgttca ggagcagctg ccctactctg tcacgctcat ctccatgccc gccaccactg 37620
 aaggcaggcg gggcccttcc gtgtccgtgg agagtgcctg ctccaaactac gccaccactg 37680
 tgcaagtgaa agagtaagta ttttggaaac ctttcagcag gggttctgtt gcaagatgt 37740
 taaatggcc tcagagggtct tagacccatca aagtctcatc cagaactccc ttatctca 37800
 tctcatatct ttctcttggta ccccaactatg ctgttaccgt acctggccct tggcacttac 37860
 tggttctctt gcccaggatc cttccatccc gatacttaag gcaagaatca ctcaccccttc 37920
 aggtgtcagg ttccaggatc tgtttgcatac ttgaaatcat ctgggtttagt tatgtgtatt 37980
 agttgtttat cttctatcccc ctccactaga atgttaattt cagaagaaac ttgtctgttt 38040
 attcagtgtc gcatgcccag ggcttggaaag agtacctggc atatgttgg agttgtatgt 38100
 ttattatctt gtcagtcagc agaatgaatg gagaatgtt ggtccatggc cccaaagaaag 38160
 ttaagacccat atccatgattt caggccagag accagatgga gaaagatgtt gtgttatct 38220
 aataccatgtt atgttgcatac tctggccgtt taccatgttaa atattgtt gttatctacc 38280
 atgtgttggc cactaggatc gtgttgcac agcaggatgtt gatatactgtt gtttggaaag 38340
 tcaggaggag ctaaggatctg ttctacaacc ttattatgtt gaggaggatgtt ggaattgtgt 38400
 tcaggaggatgtt gggagaagca ttctccaaa agttaggtgtt ttaatcatgtt ctgtatgttgg 38460
 ttgagttgttgg ccagaaaagg ggctgtttaag tataaggggc ctggattatgtt aatccacgc 38520
 agatccatgtt agatgtttaag cagcaaggatgtt ttgtgacccaa gtttacattt tagaaggatc 38580
 actggatgttgg aggttggatgtt gggagggggaa aagccataaag gtatagagac tagttggaa 38640

-continued

-continued

-continued

```

taaatgactg atgtcagcaa ccccaaaaatg atacatctga tgtaagagcc cctgttcccc 43440
aataataaca tctaaactat agacatgg aatgaacaggt gcccctaagt ttcctccctc 43500
cagggttct tggccggctc ctgaggacta cacatcccta ctccctgtt tcctcatctt 43560
caggcgcagt aacagtatct ccaagttccc tggcccccagc tccccaagg agccctgtc 43620
gttcagccgt gacatcagcc gtcagaatac ctttcgttgc tccagcaggt attcacagca 43680
gatcttccgg ccctgtgacc taatccatgg ggaggtctg gggaaagggtc tctttggca 43740
ggctatcaag gtgagcgcag gcaacaattg ctttgcctt ctgccccag tccctctgtc 43800
actgtcttc ggggatttct catcactgg ccccacccca caccatgcag gatgcccaggc 43860
ctccttcctg gctttgggtg ttgggtgtgag aggtatctt caccccccacc caggecacct 43920
aaggctcaatg ttgctgttac agtgagcttgc tggacccatgg gatccagggtt ggggtttagct 43980
gtgcctgtgg ccctccgtcc tccagtcagt ggggtttgt taggtgcctg cagacccatcg 44040
taccgggcat gctacaagggc gcacacacggg gaatggctcc tgcctccctg gtgaacagtc 44100
tcaggggacta acctctctct ttcctctctc ctccctctct tctgtgtgaga actggggagg 44160
gggggtcagggt aagacgtgtg ttcagtcatttgc ggggcagcag ggctggagag ctcacccccc 44220
atccacccatc ctccctgtgtc catgtcttgc gcaactgaccc tccctgccttcc agacttctgt 44280
tcacttcagga gactcacttc tatgccaaat gaccagagcc cctgcttggc ttggcagcat 44340
ccctccctgc ctcttccccc acttccctttt tctgggttct tgcctgtctt ctgtgcattgc 44400
ccagctctcc agggaaaggagg gtttgcctcc gtgtgagtc catgttgctc caccgtcattgc 44460
cttccacaca tgaactctgt catttcgacc cggctcattgc tgcctccaa gggatgggt 44520
ggccagctgc atagattttc tcaaacaatgtt ctccagaact tccctctggc tcagcaccat 44580
taacagtcac cctccctgtt ggtgacacac aaagccacgg gcaaaatgtt ggtcatgaaa 44640
gagttaaatc gatgtgtatgc ggagacccatgg aaaaacttttgc tgactgagggtt aagaagatgg 44700
agggggcccg ggagggttgggtt gtcaccatttgc gaagagagaa gacccatggaa ataatggctt 44760
caagagaaaa tacagtttgg aattactgtc taaaagacta agcagaaaaag agcccttagag 44820
gaatatccca ctccctctaa attacagcgt aattattttgt tcaatgaaca cttaactaaaa 44880
gcaacacaaa cagggtacaa gggatgcgtt aacaaaatggat acagggttca gaagagctct 44940
cagggttatga ggtatgttgc catgaaaaca ctccaaatgtt gtaactca atgttataat 45000
cctcacctga acggccctgtt aaggggagctt ggagggggagc tccctgtatgc ctccacactcc 45060
ttggggatttgc acatgttgc tcaatgttgc caagtttgc tcaatgttgc cttaagtgg 45120
ggacacctgtt ggtctgggtt ttcctccatgg agccacttgc ccactccac cccagtttc 45180
ccatgcgtt tcaatgttgc agggctatgtt gggccatctcc atctatgtt tggtgggtt 45240
ccctccatctt gatcttagtgc ccctgtatgtt tcaatgttgc gggatgggtt gatacagggt 45300
gtacacttgc tcaatgttgc tcaatgttgc tcaatgttgc gggatgggtt gatacagggt 45360
ccaccagctt tccctggctt ccccccacccca cagggtaaatgtt tgatgcgtt cccatggacc 45420
cccaatgttgc tcaatgttgc tcaatgttgc tcaatgttgc gggatgggtt gatacagggt 45480
gagttacatttgc agggggccac actgttgc tcaatgttgc tcaatgttgc gggatgggtt gatacagggt 45540
catatgttgc agggggccac actgttgc tcaatgttgc tcaatgttgc gggatgggtt gatacagggt 45600
ttaaaggccatgg gggggccatgg ttcctggctt cccatggacc tcaatgttgc gggatgggtt gatacagggt 45660
acataggggatgttgc tcaatgttgc tcaatgttgc gggatgggtt gatacagggt 45720
cagggttgc tcaatgttgc tcaatgttgc gggatgggtt gatacagggt 45780

```

-continued

agaaggtagtggccaa ggaatcgccct ccggatcggt gaggcccacc aacaaacactg 45840
 ccagcaggcc gagagttaggg agaggtgtga gaattgtggg cttcactggaa aggttagagac 45900
 cccttcctat gcaacttgc tgggctgggt cagcagctat tcattgagtt tgcgtgtgtc 45960
 actgaaactg acccccagccaa actgttctca gttcacagcc ctgtttcaaa agaattacac 46020
 atctctaaag gcaaaccaggg cacggacaaag gcaaactggaa gaggcaaaact gtgcgtgag 46080
 atggcctggg cttgcccatac caggttattca ggtgctgggg gccccttagac caactagagc 46140
 acctcactgc ctaggaaatc aatgaagggg aaatgagttc tagcggagcc ctgaaggatc 46200
 agaattggat aaagttctta ttggcagaga ggcaccaggaa ttgaagtgcaggagcaag 46260
 acctgggagg aaagaggaga aaatcatctta ttccactgg aaacaatga ttccaagcat 46320
 agaaataata acagctgaca agtactgagt gcccctata tgcgtggac tgggctgggg 46380
 gattaacatc catgtgcatttttattcatac atgacaacact tggtttccag ataaactggaa 46440
 ctggaaaggg acagagctgg gatccctgggc taatcagtttgc ggtcgccaaag cctgagactt 46500
 tagccactgc cttcacatc ggggtccatg aaaaatagtag tagtctggaa cagtttgggg 46560
 gtacatcaag gtcgtgtgtt tttaaagcttat ggtgtctggaa ctataggaga caaatgtaaa 46620
 agagttttttt ggttactgg cttttttgtt tttttttttt tttttttttt tttttttttt 46680
 tttttttttt ttttccctgtt tctggggctt gaatcaggaa ggaggtttttt tttttttttt 46740
 tttttttttt gtagatatttgc tttttttttt tttttttttt tttttttttt 46800
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 46860
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 46920
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 46980
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47040
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47100
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47160
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47220
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47280
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47340
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47400
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47460
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47520
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47580
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47640
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47700
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47760
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47820
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47880
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 47940
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 48000
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 48060
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 48120

-continued

tggctcccttc tacatggca aggaaaggag aaaggaagtt gctcctgaga gtggtaeag 48180
 tcagtggtgg aggcctggag aggagacata acaaacaat ttgttgacaa acattttgt 48240
 aggaaggggg agagcttaaa gtttagacag tggggaaaggt ggagtcttag aggaggtaa 48300
 tgtctgaaag acagagctag ctggagcaag aagtcaatc tctgttgagc gcaggaaagg 48360
 tccaaatggg ctcaagccag agattgggg agtggggagg agggagcagc ctggatctaa 48420
 gtaaaatggg tagaggtgga gggggctg caacggccag gttttctga agttgggac 48480
 attaggagag agctgtgagg gctttggcca gccactgtgc tagtggatgg tgaaccaaag 48540
 gatggcagg agatggcagc agggaaagcag aggaagtcca ggcttcgtt tggatttggg 48600
 acaaggggaga ggcctatgga ggcctggcc ctgttgccca gttgggttc tgaagctggg 48660
 tgggcattggc ctggtaggag agcatctatg ggccttcaatt ccagattcag ggtcttagtg 48720
 atttgctggc cctgttagct cagctcatgc ttctgttcca ggcctatgg cactctatgt 48780
 gcatcatcca ccggatctg aactcgacca actgcctcat caagttgtt gttccactg 48840
 ctctggcct ggcctccagg gtcctatctt tcctggctt cttgtcacaa aggaggctg 48900
 ctgtccctt ctggcttagag ggcagaggtt tgccttagga gtcctatctt ttcccttctt 48960
 gcttcttcca atgccttctt ctgtctctg ggagctccga gacacacaca gacataattt 49020
 caccttctt cattagcaac ctttgaata atttgattag aaggacttc agaagttgt 49080
 tgactatatg tagaaaaccc tgcattttt cctgttttg ccccatatgt gtctgtaaa 49140
 acagttcatt gtcgtacccca ttttacagtg gtggcacctg aagcctcagc ctgaggccac 49200
 cgagctatgt aatttacagg gaccatgg agaccatgt tcctccact gcccctcagc 49260
 tgggtgggtt acaatgttgtt ttgtcttact gacttgctat ctggcttctt ggtgtctac 49320
 cggctggccc tggctctgcc ctctagaccc acaccacga atcttcatc ctttccacaca 49380
 tgactgcctt gtatgtatc aaagagcttg tctccccaa gtctcccat ctactgcctc 49440
 caccttgcct ttttctgtt ttcctgggtt ctggccactg cctgaaatca ttttaggaat 49500
 aagacaggac agggaaaaac aaaagcaacc ccctgttccca cctctgatgt ccactctcca 49560
 agtccctgag ctcacccctt agggctccag tggctctgcc atgaacccac tggggctgg 49620
 gagtctgctg tgcacagata ccagaccctc agaaacacaa atgccaatgt tgcatttttt 49680
 tttttttttt tgcattttttt ttttttagatg gacttcatt ctgtttccca ggctggatgt 49740
 cagttgtca atcttggctt actgcagctt ctacccctcg ggttcttagt attgttctgc 49800
 ttcagccctcc cagtagctg gactacaggc gtgtggccacc acgccccact aatttttttt 49860
 tttttttttt tgcattttttt ttttttagatg gacttcatt ctgtttccca ggctggatgt 49920
 aactccctgac ctcaggatgt tcacccctcc tggctccca aagttctggg attacaggat 49980
 gaagccaccc tgcctggctt gactgtgtt attgtataga gctttctgtt ctgattttcc 50040
 cttgtatatac accttttctcc cccttctca gtttcttctt tgcctatgtt tcctcccaag 50100
 ggccaggatgtt gagaacatcc ccatgaatgc ctgacactgtc ttttacccca ccaggacaag 50160
 actgtggtgg tggcagactt tggctctgcc cggctctatag tggaaagagag gaaaaggccc 50220
 cccatggaga aggcacccac caagaaacgc accttgcgcg agaacgaccc caagaagcgc 50280
 tacacggtgg tggaaaccc ctactggatg gcccctgaga tgctgaacgg tgatgttgc 50340
 agccctggag gggacaccccg cagaggagg agatgtctg cccttgcatac agagccctgg 50400
 gaattccagg ggaggcctgtt gaagcttagg accggatacc cagagctgag gatatttttc 50460
 ctttgcagg tggggccctca cgatgtatgtt cctgagctca gggggctggg aactgtatcg 50520

-continued

-continued

-continued

agtattgtat gcctactgtg tgcttaggtac agttctaaac acttgggta cagcagcgaa 55320
 caaaataaaag gtgccttaccc tcataagaaca tagattctag catggatct actgtatcat 55380
 acagtagata caataagtaa actatattga stattagaat gtggcagatg ctatggaaaa 55440
 agagtcaaga caagtaaaga cgattgtca gggttaccagt tgcaattta aatatggtcg 55500
 tcagagcagg cctcactgag gtgacatgac attaagcat aaacatggag gaggaggagt 55560
 aagcctgagc tgccttaggc ttccgggca gccaagccat ttccgtggca cttaggacct 55620
 ggtgtttccg attccaccc ttgataactgc attttctcta agatatggga gggaaagttt 55680
 tctccttatg ttttaagta ttaactccag ctgtccagc cttgtttagt tgttacctaa 55740
 tctttatgc aaatataatga ggttcccgta acattatgcc catttctcac agaggacta 55800
 cttaggtgaag gagtttgctt gacgttatac aaccaggaaag tagctgagcc tagatccctt 55860
 ccacccaccc catggccctg ctcatgttcc acctgcctt aatttacctc ttttccttct 55920
 agaccagcat ttcgaaatt ggaggactcc ttggaggccc tctccctgtt cctggggag 55980
 ctgggcatecc cgctgcctgc agagctggag gagttggacc acactgtgag catgcagttac 56040
 ggcctgaccc gggactcacc tccctagccc tggcccgacc ccctgcaggg ggggtttcta 56100
 cagccagcat tgccctctg tgccccatc ctgtgtgag cagggccgtc cgggcttct 56160
 gtggatttggc ggaatgttta gaagcagaac aagccattcc tattacctcc ccaggaggca 56220
 agtgggcgcga gcaccaggaa aatgtatctc cacaggttct gggcctagt tactgtctgt 56280
 aaatccaaata ttgcctgaa agctgtgaag aaaaaaaaaa cccctggct ttggggcagg 56340
 aggaatctgt tactcgaatc cacccaggaa ctccctggca gtggatttgc ggaggcttctt 56400
 gcttacacta atcagcgtga cctggaccc tggggcagga tcccagggtt aacctgcctg 56460
 tgaactctga agtacttagt ccagctgggt gcaggaggac ttcaagtgtg tggacgaaag 56520
 aaagactgtat ggctcaaaagg gtgtaaaaaa gtcagtgtat ctccccctt ctactccaga 56580
 tccctgtctt cctggagcaaa gggttggggta gttaggtttt aagagtcctt taatatgtgg 56640
 tggaaacaggc caggagttag agaaagggtt ggcttctgtt tacctgtca ctggctctag 56700
 ccagccccagg gaccacatca atgtgagagg aagcctccac ctcatgtttt caaaacttaat 56760
 actggagact ggctgagaac ttacggacaa catccttctt gtctgaaaca aacagtccaca 56820
 agcacaggaa gaggctgggg gactagaaag aggccctgcc ctctagaaag ctcagatctt 56880
 ggcttctgtt actcataactc gggtgggctc cttagtca ctcctaaac atttgccta 56940
 aagctcgatg ggttctggag gacagtgtt cttgtcacag gccttagatc tgaggagg 57000
 gagtggggatg cttagcaatc tcttggctt ggcttcatgg caaccactgc tcacccttca 57060
 acatgcctgg tttaggcgc agcttgggtt gggaaagagggtt ggtggcagag tctcaaagct 57120
 gagatgtca gagatgtac tccctgatc gggccatctg acttctaccc cccatgttg 57180
 ctctcccaac tcattagctc ctggggcagca tccctctgtt ccacatgtgc aggtactgg 57240
 aaacctccat ctggctccc agactcttagt gaaacttca tcacaactag atttgcctt 57300
 tctaagtgtc tatgagctt caccatattt aataaattgg gaatgggtt ggggtattaa 57360
 tgcaatgtgtt ggtgggttgc ttggagcagg gggaaattgtt aagggagatg ggttgctgtt 57420
 aatattatct tatctattgg gtggatgtt gaaatattgtt catagacccatg atgagttgt 57480
 ggaccagatg tcatctctgg tcagagttt cttgtatata agactgtact tatgtgtgaa 57540
 gtttgcagaac ttgttttagg gctgagccct ggactcccaag cagcagcaca gttcagcatt 57600

-continued

gtgtggctgg ttgtttctgg gctgtcccca gcaagtgtag gagtggtggg cctgaactgg 57660
 gccattgatc agactaaata aattaagcag ttaacataac tggcaataatg gagagtgaaa 57720
 acatgattgg ctcaaggaca taaaatgaga gggctgtcta gccaccccttgc ggcctgccc 57780
 acacaaactc cccatagcag agagtttca tgccaccaag tctaaaaccc tcaagcagac 57840
 acccatgtc tctagagaat atgtacatcc caccctgagc agccccccttgc ttgcagcagg 57900
 tgtgactgac tatgacccctt tcctggcttg gcttcacat gccagctgag tcattcccta 57960
 ggagccctac cctttcatcc tctctatatg aatacttcca tagcctgggt atccctggctt 58020
 gcttccctca gtgctgggtt ccaccccttgc aatggaaaga aatgaatgca agtcacccca 58080
 ccccttgggtt ttcccttacaa gtgcttgaga ggagaagacc agtttcttgc tgcttctgca 58140
 tgtggggat gtcgtagaag agtgaccatt gggaaaggaca atgtatctg gtttagtgggg 58200
 ccttgggcac aatataaattc tgtaaaccctt aaggtttttt ctcccaggca ctctcaaaac 58260
 ttgaagaatc caacttaagg acagaatatg gttcccgaaa aaaactgtatg atctggatg 58320
 cgcattgtc gcagaaccac agagcaatgg ctgggcattgg gcagaggatc tctgggttt 58380
 cctgaggctg ataacctgtg gctgaatacc cttgctaaaaa gtccaggaga cactcccttt 58440
 ggtatctttt ctctggagt catagtatgc accttgcagg gaacttccctc agcccgaggc 58500
 tgctgcaggc agcccaatgtc cccttcctcc tctgcagttt ttcccccttgc ggctgctgca 58560
 gcaccacccc cgtcaccac cacccaaacccttgcgcact ccagcccttacaaggatc 58620
 tctagatatt cattttaaact acctccacat tggaaacaat tgctgaaggg gagaggattt 58680
 gcaatgatca accacctgtg tgggaccccttgc acacccctgt ctccctcttcaacctgaa 58740
 agattccctga tgatgataat ctggacacag aagccgggc aagggttgc agcctgtat 58800
 ctcagcactt tgggaggccttgc cagcaggatgg atcacctgtg atcaagatggt tgagaacac 58860
 ctgaccaaca tggtaacccttgc taaaataca aaaaatggcc aggtgtgggt 58920
 gcacataacctt gtaatcccttgc ctactctggaa ggctgaggca ggagaatcgc ttgaacccac 58980
 aaggcagagg ttgcagtggc gcgagatcat gccattgcac tccagccctgt gcaacaagag 59040
 cccaaacttcca tctcaaaaaaa aaaaaa 59065

<210> SEQ ID NO 4

<211> LENGTH: 265

<212> TYPE: PRT

<213> ORGANISM: Human

<400> SEQUENCE: 4

Leu Thr Glu Val Lys Val Met Arg Ser Leu Asp His Pro Asn Val Leu
 1 5 10 15

Lys Phe Ile Gly Val Leu Tyr Lys Asp Lys Lys Leu Asn Leu Leu Thr
 20 25 30

Glu Tyr Ile Glu Gly Gly Thr Leu Lys Asp Phe Leu Arg Ser Met Asp
 35 40 45

Pro Phe Pro Trp Gin Gln Lys Val Arg Phe Ala Lys Gly Ile Ala Ser
 50 55 60

Gly Met Ala Tyr Leu His Ser Met Cys Ile Ile His Arg Asp Leu Asn
 65 70 75 80

Ser His Asn Cys Leu Ile Lys Leu Asp Lys Thr Val Val Val Ala Asp
 85 90 95

Phe Gly Leu Ser Arg Leu Ile Val Glu Glu Arg Lys Arg Ala Pro Met
 100 105 110

-continued

Glu	Lys	Ala	Thr	Thr	Lys	Lys	Arg	Thr	Leu	Arg	Lys	Asn	Asp	Arg	Lys
115					120				125						
Lys	Arg	Tyr	Thr	Val	Val	Gly	Asn	Pro	Tyr	Trp	Met	Ala	Pro	Glu	Met
130				135					140						
Leu	Asn	Gly	Lys	Ser	Tyr	Asp	Glu	Thr	Val	Asp	Ile	Phe	Ser	Phe	Gly
145					150				155			160			
Ile	Val	Leu	Cys	Glu	Ile	Ile	Gly	Gln	Val	Tyr	Ala	Asp	Pro	Asp	Cys
					165				170			175			
Leu	Pro	Arg	Thr	Leu	Asp	Phe	Gly	Leu	Asn	Val	Lys	Leu	Phe	Trp	Glu
					180				185			190			
Lys	Phe	Val	Pro	Thr	Asp	Cys	Pro	Pro	Ala	Phe	Phe	Pro	Leu	Ala	Ala
					195				200			205			
Ile	Cys	Cys	Arg	Leu	Glu	Pro	Glu	Ser	Arg	Pro	Ala	Phe	Ser	Lys	Leu
					210				215			220			
Glu	Asp	Ser	Phe	Glu	Ala	Leu	Ser	Leu	Tyr	Leu	Gly	Glu	Leu	Gly	Ile
					225				230			235			240
Pro	Leu	Pro	Ala	Glu	Leu	Glu	Leu	Asp	His	Thr	Val	Ser	Met	Gln	
					245				250			255			
Tyr	Gly	Leu	Thr	Arg	Asp	Ser	Pro	Pro							
					260				265						

That which is claimed is:

1. An isolated nucleic acid molecule consisting of a ³⁰ nucleotide sequence selected from the group consisting of:
 - (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
 - (b) a nucleic acid molecule consisting of the nucleic acid sequence of SEQ ID NO:1;
 - (c) a nucleic acid molecule consisting of the nucleic acid sequence of SEQ ID NO:3; and
 - (d) a nucleotide sequence that is completely complementary to a nucleotide sequence of (a)–(c).
2. A nucleic acid vector comprising a nucleic acid molecule of claim 1.
3. A host cell containing the vector of claim 2.
4. A process for producing a polypeptide comprising culturing the host cell of claim 3 under conditions sufficient for the production of said polypeptide, and recovering the peptide from the host cell culture.

5. An isolated polynucleotide consisting of a nucleotide sequence set forth in SEQ ID NO:1.

6. An isolated polynucleotide consisting of a nucleotide sequence set forth in SEQ ID NO:3.

7. A vector according to claim 2, wherein said vector is selected from the group consisting of a plasmid, virus, and bacteriophage.

8. A vector according to claim 2, wherein said isolated nucleic acid molecule is inserted into said vector in proper orientation and correct reading frame such that the protein of SEQ ID NO:2 may be expressed by a cell transformed with said vector.

9. A vector according to claim 8, wherein said isolated nucleic acid molecule is operatively linked to a promoter sequence.

* * * * *